Journal of
Experimental Psychology

Arthur W. Melton, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

id A. Grant, Associate Editor
University of Wisconsin

William K. Estes, Associate Editor
Indiana University

Consulting Editors

man H. Anderson,
University of California, Los Angeles
ames Archer, *University of Wisconsin*
Attneave, III, *University of Oregon*
us J. Burke, *Indiana University*
s Deese, *Johns Hopkins University*
M. Fitts, *University of Michigan*
k A. Geldard, *University of Virginia*
es J. Gibson, *Cornell University*
old W. Hake, *University of Illinois*
hur L. Irion, *Tulane University*
Howard H. Kendler, *New York University*
Herschel W. Leibowitz,
International Business Machines Corporation, Bethesda, Maryland
Donald B. Lindsley,
University of California, Los Angeles
Kenneth MacCorquodale, *University of Minnesota*
Quinn McNemar, *Stanford University*
Leo Postman, *University of California, Berkeley*
L. Starling Reid, *University of Virginia*
Kenneth W. Spence, *State University of Iowa*
Delos D. Wickens, *Ohio State University*

en Orr, Acting Production Manager
Ruth P. Jagolinzer, Editorial Assistant

Volume 63, 1962

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

Second-class postage paid at Lancaster, Pa.

CONTENTS OF VOLUME 63

Abelson, R. P., and Sermat, V. **Multidimensional Scaling of Facial Expressions** 546
Adams, J. A., and Chambers, R. W. **Response to Simultaneous Stimulation of Two Sense Modalities** 198
Adams, J. A., and Creamer, L. R. **Anticipatory Timing of Continuous and Discrete Responses** 84
Alluisi, E. A. See Brainard, R. W.
Amsel, A., and Penick, E. C. **The Influence of Early Experience on the Frustration Effect** .. 167
Andrews, J. M. See Siegel, S.
Appley, M. H. See Grings, W. W.
Archer, E. J. **Concept Identification as a Function of Obviousness of Relevant and Irrelevant Information** 616
Archer, E. J. See Bulgarella, R. G.
Bacon, W. E. See Hulse, S. H.
Badia, P. See Ehrenfreund, D.
Battig, W. F., Williams, J. M., and Williams, J. G. **Transfer from Verbal-Discrimination to Paired-Associate Learning** 258
Beilin, H., and Horn, R. **Transition Probability Effects in Anagram Problem Solving** 514
Berry, R. N. **Skin Conductance Levels and Verbal Recall** 275
Besch, N. F., Thompson, V. E., and Wetzel, A. B. **Studies in Associative Interference** 342
Bilodeau, E. A. See Ryan, F. J.
Bindra, D. See Mendelson, J.
Blane, H. T. **Space Perception among Unilaterally Paralyzed Children and Adolescents** 244
Bolles, R. C. **A Psychophysical Study of Hunger in the Rat** 387
Bontrager, H. See Epstein, W.
Borresen, C. R., and Lichte, W. H. **Shape-Constancy: Dependence upon Stimulus Familiarity** 91
Botwinick, J., and Brinley, J. F. **An Analysis of Set in Relation to Reaction Time** 568
Brackbill, Y. See Little, K. B.
Brackbill, Y., Kappy, M. S., and Starr, R. H. **Magnitude of Reward and Probability Learning** 32
Brainard, R. W., Irby, T. S., Fitts, P. M., and Alluisi, E. A. **Some Variables Influencing the Rate of Gain of Information** 105
Brinley, J. F. See Botwinick, J.
Broadbent, D. E., and Gregory, M. **Donders' B- and C-Reactions and S-R Compatability**. 575
Brownstein, A. J. **Predicting Instrumental Performance from the Independent Rates of Contingent Responses in a Choice Situation** 29
Bugelski, B. R. **Presentation Time, Total Time, and Mediation in Paired-Associate Learning**. 409
Bulgarella, R. G., and Archer, E. J. **Concept Identification of Auditory Stimuli as a Function of Amount of Relevant and Irrelevant Information** 254
Buss, A. H. See Geer, J. H.
Capaldi, E. J., and Senko, M. G. **Acquisition and Transfer in Partial Reinforcement** 155
Carlin, S. See Grings, W. W.
Carlson, V. R. **Size-Constancy Judgments and Perceptual Compromise** 68
Chambers, R. W. See Adams, J. A.
Champion, R. A., and McBride, D. A. **Activity during Delay of Reinforcement in Human Learning** 589
Clark, B., and Graybiel, A. **Visual Perception of the Horizontal during Prolonged Exposure to Radial Acceleration on a Centrifuge** 294
Clark, R. E. **The Role of Drive (Time Stress) in Complex Learning: An Emphasis on Prelearning Phenomena** 57
Collier, G. See Knarr, F. A.
Collins, W. E. **Effects of Mental Set upon Vestibular Nystagmus** 191
Cotton, J. W., Jensen, G. D., and Lewis, D. J. **Spontaneous Recovery Interval as a Factor In Reacquisition of T Maze Behavior** 555
Creamer, L. R. See Adams, J. A.
Crovitz, H. F., and Daves, W. **Tendencies to Eye Movement and Perceptual Accuracy** 495
Curtin, M. E. See Estes, B. W.
Daves, W. See Crovitz, H. F.
Davis, K. G., and Hess, H. F. **The Effectiveness of Concepts at Various Levels of Awareness**. 62
Derks, P. L. **The Generality of the "Conditioning Axiom" in Human Binary Prediction** 538
Dyal, J. A. **Latent Extinction as a Function of Number and Duration of Pre-Extinction Exposures** 98

Ehrenfreund, D., and Badia, P. Response Strength as a Function of Drive Level and Pre- and Postshift Incentive Magnitude........ 468
Ellis, N. R. See Jones, R. W.
Epstein, W. A Test of Two Interpretations of the Apparent Size Effects in a Distorted Room. 124
Epstein, W., Bontrager, H., and Park, J. The Induction of Nonverdical Slant and the Perception of Shape........ 472
Erdmann, R. L. Brightness Discriminations with Constant Duration Intermittent Flashes.. 353
Erlick, D. E. The Ability to Filter Noise from a Visual Task When the Noise and Signal are Presented Sequentially........ 111
Erlick, D. E. Perception of the Most Frequent Category of a Random Series as a Function of the Number of Categories........ 115
Estes, B. W., Miller, L. B., and Curtin, M. E. Supplementary Report: Monetary Incentive and Motivation in Discrimination Learning—Sex Differences........ 320
Fehrer, E., and Raab, D. Reaction Time to Stimuli Masked by Metacontrast........ 143
Fitts, P. M. See Brainard, R. W.
Fitts, P. M., and Switzer, G. Cognitive Aspects of Information Processing: I. The Familiarity of S-R Sets and Subsets........ 321
Foster, H. The Operation of Set in a Visual Search Task........ 74
Fowler, H., and Trapold, M. A. Escape Performance as a Function of Delay of Reinforcement........ 464
Fowler, R. L., and Kimmel, H. D. Operant Conditioning of the GSR........ 563
Fuchs, A. H. The Progression-Regression Hypotheses in Perceptual-Motor Skill Learning.. 177
Gagné, R. M., and Smith, E. C., Jr. A Study of the Effects of Verbalization on Problem Solving........ 12
Games, P. A. A Factorial Analysis of Verbal Learning Tasks........ 1
Geer, J. H., and Buss, A. H. Supplementary Report: Generalization of a Nonverbal Response to Aggressive Verbal Stimuli........ 413
George, F. H. Acuity and the Statistical Theory of Figural Aftereffects........ 423
Ginsberg, R. See Suppes, P.
Glucksberg, S. The Influence of Strength of Drive on Functional Fixedness and Perceptual Recognition........ 36
Goldstein, D. A., and Newton, J. M. Transfer of Training as a Function of Task Difficulty in a Complex Control Situation........ 370
Goldstein, M. J. A Test of the Response Probability Theory of Perceptual Defense........ 23
Gonzalez, L. M. See Logan, F. A.
Gormezano, I., and Moore, J. W. Effects of Instructional Set and UCS Intensity on the Latency, Percentage, and Form of the Eyelid Response........ 487
Grant, D. A. See Hartman, T. F.
Graybiel, A. See Clark, B.
Greeno, J. G. See LaBerge, D.
Gregory, M. See Broadbent, D. E.
Grings, W. W., Carlin, S., and Appley, M. H. Set, Suggestion, and Conditioning........ 417
Hartman, T. F., and Grant, D. A. Effects of Pattern of Reinforcement and Verbal Information on Acquision, Extinction, and Spontaneous Recovery of the Eyelid CR........ 217
Hershenson, M. Reaction Time as a Measure of Intersensory Facilitation........ 289
Hess, H. F. See Davis, K. G.
Hillix, W. A. See Premack, D.
Hillner, K. See Peterson, L. R.
Hodge, M. H., and Pollack, I. Confusion Matrix Analysis of Single and Multidimensional Auditory Displays........ 129
Holden, K. B., and Rotter, J. B. Supplementary Report: A Nonverbal Measure of Extinction in Skill and Chance Situations........ 519
Horn, R. See Beilin, H.
House, B., and Zeaman, D. Reversal and Nonreversal Shifts in Discrimination Learning in Retardates........ 444
Howell, W. C. On the Heterogeneity of Stimulus and Response Elements in the Processing of Information........ 235
Hulse, S. H., and Bacon, W. E. Supplementary Report: Partial Reinforcement and Amount of Reinforcement as Determinants of Instrumental Licking Rates........ 214
Hunt, D. P. The Effect of Stimulus Similarity on the Acquision and Extinction of a Conditioned Response........ 278
Irby, T. S. See Brainard, R. W.
Jensen, G. D. See Cotton, J. W.
Jones, R. W., and Ellis, N. R. Inhibitory Potential in Rotary Pursuit Acquision by Normal and Defective Subjects........ 534
Kappy, M. S. See Brackbill, Y.
Kassel, S. H. See Little, K. B.

KIMMEL, H. D. See FOWLER, R. L.
KING, R. A. See OLSON, G.
KNARR, F. A., AND COLLIER, G. **Taste and Consummatory Activity in Amount and Gradient of Reinforcement Functions** 579
LABERGE, D., GREENO, J. G., AND PETERSON, O. F. **Nonreinforcement and Neutralization of Stimuli** 207
LAND, V. See PETERSON, L. R.
LEVINE, M. **Cue Neutralization: The Effects of Random Reinforcements upon Discrimination Learning** 438
LEWIS, D. J. See COTTON, J. W.
LICHTE, W. H. See BORRESEN, C. R.
LITTLE, K. B., BRACKBILL, Y., AND KASSEL, S. H. **A Test of a General Utility Theory for Probability Learning** 404
LOGAN, F. A., AND GONZALEZ, L. M. **Supplementary Report: Shift from Nonreward to Negatively Correlated Reward** 416
LUCE, R. D., AND SHIPLEY, E. F. **Preference Probability between Gambles as a Step Function of Event Probability** 42
MCBRIDE, D. A. See CHAMPION, R. A.
MCFARLAND, J. H., WERNER, H., AND WAPNER, S. **The Effect of Postural Factors on the Distribution of Tactual Sensitivity and the Organization of Tactual-Kinaesthetic Space** ... 148
MARSHALL, L. M. See PARDUCCI, A.
MAYZNER, M. S. **Verbal Concept Attainment: A Function of the Number of Positive and Negative Instances Presented** 314
MAYZNER, M. S., AND TRESSELT, M. E. **Anagram Solution Times: A Function of Word Transition Probabilities** 510
MECHANIC, A. **The Distribution of Recalled Items in Simultaneous Intentional and Incidental Learning** 593
MEDNICK, S. A., AND WILD, C. **Reciprocal Augmentation of Generalization and Anxiety** 621
MENDELSON, J., AND BINDRA, D. **Combination of Drive and Drug Effects** 505
MILLER, L. B. See ESTES, B. W.
MOORE, J. W. See GORMEZANO, I.
MUESER, G. E. See YNTEMA, D. B.
MUSGRAVE, B. S. **The Effect of Nonsense-Syllable Compound Stimuli on Latency in a Verbal Paired-Associate Task** 499
NEWTON, J. M. See GOLDSTEIN, D. A.
OLSON, G., AND KING, R. A. **Supplementary Report: Stimulus Generalization Gradients along a Luminosity Continuum** 414
OSLER, S. F., AND WEISS, S. R. **Studies in Concept Attainment: III. Effect of Instructions at Two Levels of Intelligence** 528
PARDUCCI, A., AND MARSHALL, L. M. **Assimilation vs. Contrast in the Anchoring of Perceptual Judgments of Weight** 426
PARK, J. See EPSTEIN, W.
PENICK, E. C. See AMSEL, A.
PETERSON, L. R., AND PETERSON, M. J. **Minimal Paired-Associate Learning** 521
PETERSON, L. R., SALTZMAN, D., HILLNER, K., AND LAND, V. **Recency and Frequency in Paired-Associate Learning** 396
PETERSON, M. J. See PETERSON, L. R.
PETERSON, O. F. See LABERGE, D.
POLLACK, I. See HODGE, M. H.
POSTMAN, L. **The Temporal Course of Proactive Inhibition for Serial Lists** 361
PREMACK, D., AND HILLIX, W. A. **Evidence for Shift Effects in the Consummatory Response**. 284
PREMACK, D., AND PUTNEY, R. T. **Relation between Intersession Interval Frequency of Competing Responses and Rate of Learning** 269
PROKASY, W. F., JR., AND WHALEY, F. L. **Manifest Anxiety Scale Score and the Ready Signal in Classical Conditioning** 119
PUTNEY, R. T. See PREMACK, D.
RAAB, D. See FEHRER, E.
RICHARDS, J. M., JR. **The Cue Additivity Principle in a Restricted Social Interaction Situation**. 452
ROSENBAUM, M. E., AND TUCKER, I. F. **The Competence of the Model and the Learning of Imitation and Nonimitation** 183
ROTTER, J. B. See HOLDEN, K. B.
RYAN, F. J., AND BILODEAU, E. A. **Countertraining of a Simple Skill with Immediate and 1-Week Delays of Informative Feedback** 19
SADLER, E. . See TEICHNER, W. H.
SALTZMAN, D. See PETERSON, L. R.
SENKO, M. G. See CAPALDI, E. J.
SERMAT, V. See ABELSON, R. P.

Shipley, E. F. See Luce, R. D.
Siegel, S., and Andrews, J. M. **Magnitude of Reinforcement and Choice Behavior in Children** 337
Simon, S. H. **Response-Mediated Generalization with Simple Skeletal-Motor Responses** 458
Slamecka, N. J. **Retention of Connected Discourse as a Function of Duration of Interpolated Learning** 480
Smith, E. C., Jr. See Gagné, R. M.
Snyder, H. L. **Saccharine Concentration and Deprivation as Determinants of Instrumental and Consummatory Response Strengths** 610
Sperling, S. E. **Extinction Effects Following Nondifferential Reinforcement of an Irrelevant Stimulus** 50
Stabler, J. R. **Performance in Instrumental Conditioning as a Joint Function of Time of Deprivation and Sucrose Concentration** 248
Starr, R. H. See Brackbill, Y.
Sundland, D. M., and Wickens, D. D. **Context Factors in Paired-Associate Learning and Recall** 302
Suppes, P., and Ginsberg, R. **Application of a Stimulus Sampling Model to Children's Concept Formation with and without Overt Correction Responses** 330
Switzer, G. See Fitts, P. M.
Teichner, W. H., and Sadler, E. **Effects of Exposure Time and Density on Visual Symbol Identification** 376
Thompson, R. W. See Trabasso, T. R.
Thompson, V. E. See Besch, N. F.
Trabasso, T. R., and Thompson, R. W. **Supplementary Report: Shock Intensity and Unconditioned Responding in a Shuttle Box** 215
Trapold, M. A. See Fowler, H.
Tresselt, M. E. See Mayzner, M. S.
Tucker, I. F. See Rosenbaum, M. E.
Tyler, V. O., Jr. **Sensory Integration with and without Reinforcement** 381
Uhr, L., Vossler, C., and Uleman, J. **Pattern Recognition over Distortions by Human Subjects and by a Computer Simulation of a Model for Human Form Perception** 227
Uleman, J. See Uhr, L.
Vossler, C. See Uhr, L.
Wapner, S. See McFarland, J. H.
Weiss, S. R. See Osler, S. F.
Werner, H. See McFarland, J. H.
Wetzel, A. B. See Besch, N. F.
Whaley, F. L. See Prokasy, W. F., Jr.
Wickens, D. D. See Sundland, D.M.
Wild, C. See Mednick, S. A.
Williams, J. G. See Battig, W. F.
Williams, J. M. See Battig, W. F.
Wist, E. R. **Amount, Delay, and Position of Delay of Reinforcement as Parameters of Runway Performance** 160
Wright, J. C. **Consistency and Complexity of Response Sequences as a Function of Schedules of Noncontingent Reward** 601
Yntema, D. B., and Mueser, G. E. **Keeping Track of Variables That Have Few or Many States** 391
Young, R. K. **Tests of Three Hypotheses about the Effective Stimulus in Serial Learning** 307
Zeaman, D. See House, B.

Journal of
Experimental Psychology

Vol. 63, No. 1 January 1962

A FACTORIAL ANALYSIS OF VERBAL LEARNING TASKS[1]

PAUL A. GAMES

Ohio University

The methods of factor analysis have been applied recently to measures typically associated with the experimental laboratory in order to identify sources of common variance on such tasks and to trace the changes in factorial structure associated with practice. Fleishman (1957, 1960) and Fleishman and Hempel (1954, 1955) have demonstrated, in the psychomotor area, that the factorial composition of a task may substantially change with practice. This study is an attempt to determine the extent of common variance among some verbal learning tasks and to identify the common factors that the tasks are measuring. More specifically, the type of task (paired-associate or serial list), method of presentation and response, and order of presentation were varied to see if these were related to the abilities measured. Also investigated was the extent to which the contributions of the ability factors change with early practice on verbal learning tasks.

The identification of sources of variance is greatly facilitated if reference tests are used to define factors whose interpretation, from the accumulated results of previous studies, is relatively clear. One factor clearly related to verbal learning is "Rote Memory." This factor is defined (Christal, 1958) by ". . . tests requiring the formation of associations . . ." (p. 1) and has been reported in at least 17 factor analysis studies (French, 1951, p. 219). The factor appears to transcend the "content" of the stimulus-response pairs "to be associated" since tasks involving numbers, forms, words, letters, pictures, colors, and names have all appeared on the factor. The factor has been defined by both recall and recognition response measures, and with both paired associates and single lists (of items) as the material to be learned. However, the rote memory group tests have typically utilized a present-recall task (present pair 1 . . . present 10; recall response to pair 1 . . . recall 10) rather than the anticipation method commonly used in experimental studies of verbal learning.

---

[1] Based on work submitted in partial fulfillment of the requirements for the PhD degree at the State University of Iowa. The author is indebted to Harold P. Bechtoldt for advice and assistance during the course of the investigation.

The serial anticipation task of verbal learning clearly involves the formation of associations. However, in terms of formal task similarity, such tasks have no counterpart in tests of Rote Memory, but are similar (on the early trials) to Span Memory tests. Span tests require the accurate reproduction of temporal sequences of stimuli. A Span Memory factor has been rather clearly identified in at least two studies (Brener, 1940; Woodrow, 1939) and suggested by the results of others (French, 1951, p. 240); however, sufficient measures to define both the Rote Memory and Span Memory factors had never been included in the same analysis, until Kelley (1954) rather clearly demonstrated the linear independence of the

TABLE 1

THE TEST BATTERY

Test No.	Test Name and Mode of Presentation	Stimulus Exposure Time (Sec/ Stimulus)	Response Time (Sec/ Stimulus)	Time between Trials (Sec.)	Total Time (Approx. Minutes)
	Day 1: Group Tests				
0′	Peripheral Span (flash cards)	.5	a	NA	3
1	Letter Span (flash cards)	2	a	NA	10
2	Symbol-Numbers (film)	4	6	8	10
3	Repeated Span (film)	1	20b	2	6
4	Word Memory (auditory)	2	4	NA	7
5	Random Pairs—Present-Recall (film)	2	2	4	7
	Free Recall				1
	Day 2: Individual Tests				
A	Warm-up Task (Missouri drum)	2	2	0	4
6	Serial Task-Anticipation (Missouri drum)	2	2	6	6
	Free Recall				1
7	Random Pairs—Present-Recall (Missouri drum)	2	2	6	6
8	Random Pairs—Anticipation (Hull drum)	2	2	4	6
	Free Recall				1
9	Repeated Span (Missouri drum)	2	20b	0	5
10	Constant Pairs—Anticipation (Hull drum)	2	2	4	4
	Free Recall				1
11	Constant Pair—Present-Free Recall (Missouri drum)	2	20b	0	6
	Day 3: Group Tests				
0″	Peripheral Span (flash cards)	.5	a	NA	3
12	S-2 Letter Span (film)	1.5	a	NA	8
13	R-3 First Names (flash cards)	4	6	NA	10
14	Constant Pairs: Present-Recall (film)	2	2	4	7
	Free Recall				1
15	Number Span (auditory)	1	a	NA	6
16	Random Serial Pairs: Present-Recall (film)	2	2	4	7
	Free Recall				1

Note.—NA = not applicable; a = see text; b = total response time allowed, no separate stimuli presented.

two factors. The early trials on a serial task thus might show a high loading on the Span factor, while the later trials shift toward the Rote Memory factor as new associations are formed.

In addition to the present-recall vs. anticipation method differentiation, verbal learning tasks vary in terms of the constancy of the sequence of presentation. Serial tasks, of necessity, have a constant order over all trials; whereas paired-associate tasks may or may not have a constant order. Another differentiation between the serial and paired-associate tasks is that in the former, each item serves as both stimulus and response. Any or all of these differentiations may affect the loadings on the Rote Memory or Span Memory factors, or may be associated with systematic individual differences, and hence may define separate factors. To investigate these possibilities and the changes in factor loadings associated with practice, the following specific hypotheses were formulated, and the battery of 17 tests listed in Table 1 was developed to investigate these hypotheses.

Hypotheses Tested

Prerequisite hypothesis.—Hypothesis 1 was that four of the tests used by Kelley (1954) and two tests used by other investigators (French, 1951) define two linearly independent factors—Rote Memory (Tests 2, 4, and 13) and Span Memory (Tests 1, 12, and 15). These measures are designated the "reference tests." All other measures are referred to as the "experimental tests." This hypothesis was labeled prerequisite in the sense that the following hypotheses, as stated, could be tested only if this hypothesis were retained. This hypothesis was tested by Part I of the analysis.

Major hypothesis.—Hypothesis 2 was that there are positive correlations between the reference and experimental tests and these correlations are accounted for by the two reference factors, Rote Memory and Span Memory. This hypothesis was tested, using the correlations between the reference and experimental tasks, (*a*) by the magnitude of the observed correlations, and (*b*) by the insignificance of residuals after the effects of the two reference factors have been eliminated.

Supplemental hypotheses.—One or more common factors *restricted to the experimental tasks* may or may not appear even if the major hypothesis is retained. Such factors could be compatible with the major hypothesis. The following possible factors were suggested by the preceding analysis. Tests that would define such factors were represented in the design of the study.

Hypothesis 3a was that a group factor is defined by the anticipation method tasks, (Tests 6, 8, and 10). Hypothesis 3b was that a group factor is defined by the constant order tasks (Tests 3, 6, 9, 10, 11, and 14). Hypothesis 3c was that a group factor is defined by the repeated Span and Serial tasks (Tests 3, 6, and 9). These hypotheses were tested by Part II of the analysis.

Special hypotheses.—Hypothesis 4a was that scores on successive trials on the serial task systematically shift, with practice, from the Span factor to the Rote Memory factor. Hypothesis 4b was that free recall scores on the experimental tasks have higher loadings on the Span factor than the regular anticipation or present-recall scores. These hypotheses were tested by Part III of the analysis. The specific manner in which these hypotheses were tested is given in the procedure below.

METHOD

Subjects

The sample consisted of 100 volunteers from the elementary psychology course at the State University of Iowa. To prevent conflicts in scheduling with another verbal learning experiment and to insure that no *S*s had prior experience on verbal learning tasks, *S*s were selected from those having scores between Percentiles 20 and 80 on the aIow Picture Interpretation Test. No significant differences were found between these upper and lower groups on the verbal paired-associate task in this other experiment. Thus, there is no indication that this selection affected the factorial results.

Procedure

The test battery shown in Table 1 was administered in three sessions held within 1 week. The groups in Sessions 1 and 3 ranged in size from 6 to 20 *S*s. Session 2 was an individual session. All tests were administered by *E*.

Group tests of auditory presentation were administered by means of a tape recorder placed in the front of the room. Film strip tests were given by use of a Soundview automatic projector and two Hunter-Brown timers, and were projected on a screen at the front of the room. These timers were left on throughout all parts of the group session (except when the auditory tests were given) to provide for the timing of the flash card tests. Written responses were required on all group tests. Oral responses were obtained on the individual tests administered by means of a Hull-type and a Missouri-type memory drum.

Test Battery

Table 1 lists the tests, the order and mode of presentation of the tests, the stimulus exposure time, the time allowed for a response, and the time between trials when applicable.

Warm-up tasks.—The warm-up task for both group sessions was a peripheral Span test in which a sequence of numbers was presented on a single flash card. Seven sequences of random numbers ranging from 4 to 10 digits in length were presented in order of increasing length. The number of completely correct sequences in both tasks (0′ and 0″) constituted Variable 0.

The warm-up task for the individual session, designed to familiarize *S*s with the anticipation method, consisted of four trials on a list of five pairs of typewriter symbols and numbers presented in a constant order. If the subject was not responding according to the instructions by the end of Trial 4, he was given two more trials.

Span Memory reference tests.—The three Span Memory reference tests allowed 10 sec. for response at the end of sequences of five items or less, 15 sec. for sequences of six to eight items, and 20 sec. for sequences over eight items. The three reference tests are slight modifications of Kelley's (1954) Span tests.

Test 1: Letter Span. After seeing a sequence of letters one at a time, *S* must write the sequence. Two test series of six items each were presented, the items ranging from 5 to 10 letters in length. Within each test series the six items were arranged in increasing order of length. Two sequences of 4 letters in length preceded the first test sequence. The score was the number of sequences completely correct.

Test 12: Letter Span. Same as Test 1 except presentation of letter sequences was done by film strip rather than flash cards.

Test 15: Number-Span. Same as Test 1 except presentation of number sequences was by a tape recorder.

Rote Memory reference tests.—The score on each of these tests is the number of correct responses.

Test 2: Symbol-Numbers. Ten pairs of simple figures were associated with two-digit numbers. Each of the three trials consisted of a presentation of all pairs and a test of each stimulus alone. Preceding the test, *S*s were given a practice task of four pairs to familiarize them with the procedure.

Test 4: Word Memory. This is Kelley's (1954) "Memory for Words I." Pairs of unrelated one-syllable nouns were read aloud twice, then the first word of each pair was read. The 20-item test was given in two parts, administered consecutively with 10 pairs per part.

Test 13: First Names. This is a task of associating first and last names. The test consisted of a learning presentation of a list of 10 pairs of first and last names, and a test presentation of the last names only, followed by another presentation of both names and another test presentation. A second list of 10 names was then given with two learning and two test presentations as above.

Experimental tests.—To reduce the possibility of specific content factors that might prevent a test of the specified hypotheses, all experimental tests used only the 21 consonant letters as stimuli and responses. Letters were chosen because of high initial familiarity and response availability that should increase rate of learning and that might reduce any "learning to learn" effects. All serial and span tests consisted of a list of 10 consonants while paired-associate tasks involved 10 pairs, or 20 consonants. Estimates of initial associational values of pairs of consonants have been based upon data from an unpublished study by Bechtoldt (1956) in which 90 *S*s gave two single-letter responses to each consonant. Only one association that had received a total frequency over seven in the above study has been used as a correct association in the lists for this experiment. The *E* also attempted to avoid any associations that might have been learned in previous tasks in this study or acquired from experiences outside the laboratory; (e.g., PT as in PT boats, etc.). (Test 11 was added to the experimental battery last; no list was found that could meet the criterion listed, so several pairs of Test 5 were reversed and used.)

Since the correct associational strength has been kept low, but high associations of each stimulus with other incorrect letters remain, these tests may be conceived of as low association-high competition tasks in contrast to the reference tests that are best described as low association-low competition tasks.

Six trials were given on most experimental tasks with three scores obtained from each set of trials; the three scores were defined as the total number of correct responses on successive blocks of two trials, i.e., Trials 1 and 2, 3 and 4, and 5 and 6. These three block scores were designated the a, b, and c measures, respectively. Free recall tests, in which *S* was asked to write down, or recite as many S-R pairs as he could recall, were designated as the d measure when given. The number of trials and the usage of a final free recall test can be determined for each test by consulting Table 3. (Trials 5 and 6 of Test 10 were run, but dropped from the analysis because of an extreme ceiling effect.) All group experimental tests were given by film strip; all individual tests, by memory drum.

Repeated Span and Serial tests.[2]—Test 3: Repeated Span. Each of the four trials consisted of a ready signal followed by the separate exposure of each of 10 consonants with 20 sec. allowed at the end of the list for *S*s to write the sequence.

Test 9: Repeated Span. Same as Test 3 with 10 blanks on the memory drum giving *S* 20 sec. to verbally recite the sequence. The last blank contained an asterisk as the ready signal for the coming trial.

Test 6: Serial Task. The one presentation trial was followed by six test trials using the anticipation method. An asterisk after the last letter indicated the start of a new trial and served as the cue for the anticipation of the first letter of the list.

Experimental Letter-Pair tasks.—Since most of these tests differ in only two ways (constant order of pairs and test stimuli over trials vs. random order; and method of presenting and testing), they may be described by title alone. The anticipation method tasks all involved one presentation trial prior to the prescribed number of test trials; a present-recall trial consisted of a presentation series followed by a recall series with each stimulus presented alone. The present-free recall task differs from the above only in that the recall series was an interval in which *S* attempted to write down as many pairs as he could.

Test 5: Random Pairs; Present-Recall (film strip)

Test 7: Random Pairs; Present-Recall (Missouri memory drum)

Test 8: Random Pairs; Anticipation (Hull memory drum)

Test 10: Constant Pairs; Anticipation (Hull memory drum)

Test 11: Constant Pairs; Present Free Recall (film strip)

Test 14: Constant Pairs; Present Recall (film strip)

Test 16: Random Serial Pairs; Present Recall (film strip). These pairs were constructed by taking adjacent letters on a serial list of 10 letters as pairs, plus the tenth letter and the first letter as a pair.

Analysis of Data

Product-moment correlation coefficients were computed for all pairs of measures using IBM punched card procedures. The analysis of the correlations involved three parts:

1. The factoring of the reference tests (1, 2, 4, 12, 13, 15).

2. The projecting of the second-block scores of the experimental tests (3b, 5b, 6b, 7b, 8b, 9b, 10b, 11b, 14b, 16b) upon the factors defined in Part 1, and the factoring of the residual matrix of these measures.

3. The projection of the free association trials and first and third blocks of the experimental tests onto the factors defined in Parts 1 and 2 (3a, 5a, 5c, 6a, 6c, 6d, 7a, 7c, 8a, 8c, 8d, 9a, 10a, 11a, 11c, 14a, 14d, 16a, 16d, 0). Table 2 designates submatrices of the complete correlation matrix according to this division.

---

[2] On Tests 3, 9, and the free recall measure of Test 6, the number of correct responses was defined as the number of consonants reproduced on the correct spaces of the answer sheet; however, if a group of three or more consonants in correct order was incorrectly located, the group was scored as the number of correct letters minus the number of spaces the group was displaced.

TABLE 2

THE COMPLETE CORRELATION SUPERMATRIX

Measures	Reference Tests	Experimental Tests	Supplementary Measures
Reference tests	1	2	4
Experimental tests (b measures only)	2′	3	5
Supplementary measures of experimental tests (a, c, and d measures)	4′	5′	6

The projection procedure in Parts 2 and 3 was the following variation of the multiple-group factoring technique originated by Bechtoldt (1956):

Let R_1 = the matrix of intercorrelations of the "reference" variables defining the factor space; the set of "reference" variables for Part 3 will include the reference tests of Part 1 and the variables of Part 2.

Let R_2 = the matrix of correlations of the "variables to be projected" with the appropriate set of "reference" variables.

Let E = the weighing matrix with entries of +1 or −1 as determined by the centroid factoring of R_1;

Define and compute:

$T = ER_1E'$, a symmetric matrix

C, a triangular matrix, where $CC' = T$ and C is defined by the diagonal factoring steps of the multiple group method of factoring.

$(C')^{-1}$ = the inverse of the transpose of C, Then $F_2 = R_2E'(C')^{-1}$.

The entries of F_2 are the factor loadings of the "projected" variables on the orthogonal centroid factors defined by the "reference" variables.

1. The centroid factoring of the Part 1 or reference variables (Matrix 1) yielded two factors (I and II) significant by a criterion based on five approximation procedures.[3] The third factor extracted was rejected by four of five tests, therefore only two factors were retained. One cycle of multiple group factoring was necessary to stabilize each communality to within ±.03 of its value the previous cycle.

2. The loadings of the experimental tests on Factors I and II were determined by the projection method using the data of Matrix 2. The matrix of two factor residuals of the intercorrelations of the experimental tests, Matrix 3, was computed using the loadings on Factors I and II, by the usual *R-FF′* procedure. Centroid factoring of the residuals yielded three additional factors (III, IV, and V.) Factor V was significant on three of the five approximate tests and therefore was retained. Three cycles of the multiple group method were necessary to stabilize the communalities.

3. The loadings of the remaining measures on Factors I and II were determined by the data of Matrix 4, and the loadings on Factors III, IV, and V were determined by the correlations of Matrix 5. In the latter matrix some of the correlations are intertrial correlations on the same task and therefore include "specific factor" variance (e.g., the correlation of 3a with 3b). Since this specific variance is to be excluded from the analysis it is necessary to derive stable estimates of the intertrial correlations based on the common factors of the battery. This estimate was secured by three successive cycles of the multiple group method.

To test the hypotheses previously specified, Factors I and II were rotated only with respect to each other, i.e., they were kept orthogonal to the set of Factors III, IV, and V defined by the experimental tests alone. Similarly Factors III, IV, and V were rotated with respect to each other only. Within these limitations the factors were graphically rotated to an oblique positive simple structure in accordance with Thurstone's (1947) criterion. Loadings on all tests were used in determining the location of the factors.

RESULTS AND INTERPRETATIONS

The factor loadings of all tests on the five rotated factors, the portion of the variance accounted for by the two reference factors, and the mean and variance of the raw scores for each measure are given in Table 3.[4]

---

[3] The criterion of significance adopted consisted of the application of five approximation tests of significance—first those of Mosier, Guilford and Lacy, and McNemar; if all yielded the same result, no further tests were run. If one of the tests differed from the other two, two additional tests were run, Tucker's and Burt's (Thompson, 1951, pp. 121–123).

[4] The complete orthogonal *F* matrix, the transformation matrix, and the product-moment correlations, and fifth factor residuals for all variables have been deposited with the American Documentation Institute. Order Document No. 6905 from ADI Auxiliary

TABLE 3

Rotated Factor Loadings, Communalities on the Reference Factors, Means, and Variances of the Variables

Test No.	Test Name and Mode of Presentation	Factors					Reference Factor Communality	Mean	σ^2
		SM(I)	RM(II)	III	IV	V			
2	Letter Span (cards)	72	04	−03	03	09	60	4.40	4.18
12	Letter Span (film)	76	12	−09	−04	−05	72	4.28	4.61
15	Number Span (auditory)	60	04	12	04	−04	43	4.43	5.51
0	Peripheral Span (cards)	63	−03	−02	08	22	42	8.05	3.69
2	Symbol-Numbers (film)	−06	69	−11	−03	10	51	13.66	36.58
4	Word Memory (auditory)	−02	72	05	04	−04	57	12.07	23.17
13	First Names (cards)	03	70	05	00	−07	55	24.48	65.51
3a	Repeated Span (film)	60	−10	−12	−15	28	36	9.92	11.09
3b	Same	42	27	−04	04	59	35	15.04	14.90
9a	Repeated Span (drum)	48	−10	12	31	25	19	9.36	13.57
9b	Same	30	29	05	19	30	26	14.07	23.27
5a	Random Pairs: Present-Recall (film)	28	42	30	20	−02	38	5.55	12.75
5b	Same	11	59	30	−10	23	45	11.51	29.01
5c	Same	22	56	36	−05	19	49	14.22	28.19
6a	Serial Task: Anticipation (drum)	−05	58	−04	35	07	36	5.23	12.68
6b	Same	01	48	01	46	−02	26	9.66	20.72
6c	Same	−06	54	08	56	06	31	12.38	22.54
6d	Same	−04	60	−04	43	−01	38	6.88	7.41
7a	Random Pairs: Present-Recall (drum)	20	45	33	13	−02	33	5.38	12.20
7b	Same	26	58	39	01	−02	54	10.40	21.78
7c	Same	12	64	42	06	12	52	12.79	20.91
8a	Random Pairs: Anticipation (drum)	24	42	56	16	−01	33	3.44	9.39
8b	Same	19	40	61	18	02	27	7.53	21.55
8c	Same	21	47	57	25	05	35	10.81	28.77
8d	Same	16	44	54	24	04	29	5.26	5.35
10a	Constant Pairs: Anticipation (drum)	32	34	45	22	04	32	5.32	11.24
10b	Same	21	47	45	30	16	26	10.92	22.13
10d	Same	21	52	26	23	18	42	5.32	4.86
11a	Constant Pairs: Present-Free Recall (drum)	14	38	16	04	30	22	7.03	6.31
11b	Same	04	53	22	16	35	33	11.25	9.81
11c	Same	11	41	28	07	23	22	13.60	12.70
14a	Constant Pairs: Present Recall (film)	24	38	51	08	08	30	9.98	14.00
14b	Same	05	44	40	08	19	23	16.81	13.59
14d	Same	14	29	37	15	09	15	8.03	3.79
16a	Random Serial Pairs: Present-Recall (film)	08	32	34	05	12	14	5.13	6.33
16b	Same	08	51	54	−05	−07	33	9.70	20.45
16c	Same	08	50	44	−05	04	30	11.23	24.72
16d	Same	09	45	35	13	23	26	5.64	5.31

Note.—Decimals omitted from factor loadings and communalities.

Prerequisite hypothesis.—Hypothesis 1 was that the reference tests define two linearly independent factors, Rote Memory (RM) and Span Memory (SM). This hypothesis was retained. The Span reference tests displayed high loadings on the first factor and the Rote Memory reference variables were highest on the second factor. The correlation between the Rote Memory and Span Memory factors was .32.

High correlations were found between the Peripheral Span task and the temporal span tasks, so that the ensuing projection of Task O on the Span factor was as high as that of one reference variable. Although the mean score of the Peripheral Span task was considerably higher than that of the temporal span tests, the factorial composition was essentially the same.

Major hypothesis.—Hypothesis 2 was that there are positive correlations between the reference and experimental tests, and these correlations are accounted for by the two reference factors. This hypothesis was retained. All but one of the 186 correlations between the experimental

---

Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

and reference tasks were positive and 108 of these correlations were above .254, the *r* corresponding to the 1% level of significance. The adequacy of the two reference factors was demonstrated by the lack of significance of the two factor residuals of Matrix 2. All experimental variables showed loadings of .29 or higher on one or both of the reference factors.

The Letter Pair tests all displayed their maximal loadings on the Rote Memory factor, as expected, and had very low loadings on the Span factor. However, the median loading of the Rote Memory reference tests was .70, while the median loading of all Letter Pair measures was .47. This decrease may have been partly due to the low association-low competition nature of the reference tests and the contrasting low association-high competition character of the Letter Pair tasks.

The median loading of the Letter Pair tests on the Span factor was .20. While the factor loadings of the Letter Pair tests were relatively low, the values were certainly not centered around the expected zero value. One possible explanation of this result may lie in the composition of the reference tests. The Span reference tests consisted of two Letter Span tests presented visually and only one Number Span test with auditory presentation. Hence, any "visual letter discrimination or general letter content" source of test variance was common to the two Span reference tests, and was, therefore, included in that factor, while other "content" characteristics were not. The fact that the Letter Span tests had communalities of .70 and .62 in the two-factor reference space while the Number Span and Peripheral Number Span tests had communalities of .43 and .42, respectively, support the interpretation of a slight bias of the Span factor toward letter content. Such a content bias would very easily account for the consistent, but low positive loadings of the Letter Pair tests.

Supplemental hypotheses.—Three common factors (III, IV, and V) restricted to the experimental tasks were extracted and left orthogonal to the reference factors. None of the reference tests had loadings above .12 on any of these factors. It will be recalled that the significance of Factor V (three of five approximate tests yielded significant values) and its retention was a somewhat arbitrary decision. Therefore two rotations were made, one with Factors III, IV, and V the other with only Factors III and IV. The loadings on Factor III were about the same in both solutions, the largest difference being .08 and the mean of the absolute differences being only .018. The loadings on Factor III for the five-factor solution are found in Table 3.

All Letter Pair tasks had loadings above .20 on Factor III. The only zero loadings were for the reference tests, the serial task and the repeated Span tasks. The factor was therefore specific to letter pair tasks, and was designated as a "Letter Pairs Content Factor." It is suggested, on the basis of *S*s' verbal reports, that the factor represents a fluency in making associations between letter pairs. This interpretation is supported by the apparent reduction in loadings accompanying the use of the free recall procedure, since if the stimulus is not presented to elicit the created association, the variance associated with this factor should be reduced. No simple formulation, however, seems to account for the other differential loadings on this factor, and any complex formulation on the basis of these limited data would be of questionable value.

Factor IV failed to account for much of the variance of the test battery. It also failed to demonstrate the stability of Factor III in the two rotational solutions; its loadings changed appreciably. The five-factor solution suggested an anticipation method factor as in Hypothesis 3a, since 9 of the 10 loadings above .20 were on anticipation method tasks (6, 8, and 10), the only exception being a loading of .31 on the first block of the individual Repeated Span task (9a). The four-factor solution, however, is not as readily interpretated, the loadings above .20 on Factor IV were: Test 3b, .47; Test 9a, .38; Test 9b, .36; Test 6a, .29; Test 6c, .40; Test 6d, .27; Test 10d, .28; Test 11a, .25; Test 11b, .37; Test 11c, .21; Test 16d, .25; Test 0, .22.

The presence of loadings on Tests 3, 9, and 6 suggests that a group factor is defined by the Repeated Span and Serial Tasks (Hypothesis 3c), but the other loadings suggest a greater generality. Hypothesis 3b, that a factor is defined by the constant-order tasks, is supported by the fact that Tests 3, 6, 9, and 11 all have a constant presentation order. However if this is the crucial differentiating factor, where are the a, b, and c measures of Tests 10 and 14, which also involve a constant order? Thus no well supported statement may be made about Hypotheses 3a, 3b, and 3c, except that at least one of the three is false. Which, if any, is true will have to be investigated in future studies.

Factor V, in the five-factor solution also accounted for relatively little test variance, and had only one measure with a large loading; the second measure of the group Repeated Span test (3b) had a loading of .59. The other repeated span measures and the Constant Pairs-Free Recall measures all had loadings between .23 and .35, suggesting that the free recall mode of response might define the factor. However, the free recall trials of the other tests failed to show substantial loadings on Factor V; these loadings were −.01 (6d), .04 (8d), .09 (14d), and .23 (16d). Thus no interpretation seemed reasonably supported.

Special hypotheses.—Hypothesis 4a stated that scores on the serial learning task shift from the Span factor to the Rote Memory factor with practice. This hypothesis was rejected; the Serial task displayed high Rote Memory loadings on all blocks of trials and approximately zero loadings on the Span factor on all blocks. Hypothesis 4b, that free recall scores on the experimental tasks have higher loadings on the Span factor than the regular anticipation or present recall scores was also rejected.

The Repeated Span tests (3 and 9), however, did exhibit a consistent change in factor loadings with practice. The loadings on these tests shifted from the Span factor toward the Rote Memory factor as the amount of practice was increased. These measures clearly changed, with practice, from relatively pure span measures to a composite of the two factors. This change of factorial structure with practice was the only large consistent change observed with respect to the reference factors. There was, however, a suggestion of a small change in the Letter Pair tests. As practice increased from the trials of Block 1 to those of Block 3, the mean loadings decreased on the Span factor (from .21 to .15) and increased on the Rote Memory factor (from .39 to .52); at the same time, a small increase in the two factor communalities occurred (from .29 to .38). The individual tests, however, did not all demonstrate this trend, so that the most parsimonious explanation is probably to attribute the Letter Pair tests shifts in factor loadings to chance, unless this finding is replicated in another study.

DISCUSSION

Christal (1958) offers two alternative hypotheses about the Rote Memory factor. "The first is that associative memory is a rather limited and artificial factor which reflects the extent to which examinees resort to use of 'memory crutches' in a formal testing situation" (p. 22). According to this interpretation, there should be a systematic reduction in Rote Memory loadings for free recall measures, since the absence of the initial stimulus portion of the "memory crutch" should greatly reduce the usefulness of this device. No such reduction was found.

"An alternative hypotheses . . . is that it [Rote Memory] represents ability to reproduce stimuli after limited exposure. This ability may have little relationship to the ability to reproduce the same materials after they have been well integrated through many exposures under various conditions and various contexts" (p. 22). The data from this study are compatable with this hypothesis as long as six trials on a verbal learning task are considered as a limited exposure. This hypothesis clearly implies that as learning progresses, loadings will shift away from the Rote Memory factor, but when this would occur is unspecified. No decrease in Rote Memory loadings from Trials 1 to 6 was observed in this experiment.

In either case, Rote Memory would seem to be the major individual difference learning parameter needed in theories of behavior in verbal learning situations. Approximately one-fifth to one-fourth of the total variance of the Letter Pair tasks and nearly one-third of the total variance of the Serial Task can be attributed to the Rote Memory factor.

Only one set of tests, the Repeated Span tests, has clearly demonstrated a need for both a Span Memory individual difference parameter and a Rote Memory parameter; the former accounts largely for the initial scores, and the latter appears to be more important as learning progresses. For theoretical purposes, it is fortunate that the more frequently used verbal learning tasks probably do not require a Span parameter. Such a factor might appear on tasks with lists shorter than the 10 item lists used in this experiment, but it is unlikely that it would appear when longer lists are used. Equally encouraging to the theoretical formulations is the fact that variations in method of presentation and the constancy of order have failed to demonstrate any consistent systematic effects on loadings on either the Rote Memory or Span Memory factors.

One practical implication of this study is that the selection of slow or fast learning *S*s can be accomplished in group sessions, even in experiments that will later use individual sessions and conventional memory drum presentation. The experiment further suggests that, at least as far as individual differences are concerned, present-recall verbal learning studies may be carried out using group testing, at a saving of time and expense, and little or no loss of generality and comparability of data.

SUMMARY

One hundred college students were given a test battery designed to investigate the loadings of a series of verbal learning tasks on the Rote Memory and Span Memory factors. The method of presentation (anticipation or present-recall), the constancy of order, and the type of task (serial or paired-associate) were varied on different tasks using consonant letters as stimuli and responses. The effects of practice over six trials and of the free recall mode of response upon the loadings on these two factors were also investigated. A set of reference factors and a set of specific factors were extracted by factor analysis. Rotations to simple struc-

ture were made within each set, but the two sets were kept orthogonal to each other.

The Rote and Span Memory factors were clearly defined as specified. The experimental verbal learning tasks displayed substantial loadings on the Rote Memory factor, and negligible or low loadings on the Span factor. Variations in method of presentation and constancy of order had no consistent systematic effect on loadings on either factor. The only substantial change of factorial composition attributable to practice was a shift of the Repeated Span tests from the Span factor toward the Rote Memory factor. The free recall mode of response had no observable effect on loadings on these factors.

Two factors (III and IV) specific to the experimental tasks were rather clearly indicated and a possible third specific factor (V) was extracted. Factor III was interpreted as a "Letter Pairs" content factor. The interpretation of Factor IV depends upon whether Factor V is retained or rejected. If it is retained, Factor IV is suggestive of an "Anticipation Method" factor; if Factor V is rejected, Factor IV's interpretation is not clear. No interpretation was offered for Factor V, if retained.

REFERENCES

BECHTOLDT, H. P. *A factor analysis of test-retest data.* Paper read at Midwestern Psychological Association, St. Louis, Missouri, May 1956.

BRENER, R. An experimental investigation of memory span. *J. exp. Psychol.*, 1940, 26, 467–482.

CHRISTAL, R. E. Factor analytic study of visual memory. *Psychol. Monogr.*, 1958, 72, (13, Whole No. 466).

FLEISHMAN, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *J. appl. Psychol.*, 1957, 4, 263–272.

FLEISHMAN, E. A. Abilities at different stages of practice in rotary pursuit performance. *J. exp. Psychol.*, 1960, 60, 126–171.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. Changes in factor structure of a complex psychomotor task as a function of practice. *Psychometrika*, 1954, 19, 239–252.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, 49, 301–310.

FRENCH, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monograph, No. 5.* Chicago: Univer. Chicago Press, 1951.

KELLEY, H. P. *A factor analysis of memory ability.* Princeton: Princeton University and Educational Testing Service, April 1954.

THOMPSON, G. *The factorial analysis of human ability.* (5th ed.) New York: Houghton Mifflin, 1951.

THURSTONE, L. L. *Multiple factor analysis.* Chicago: Univer. Chicago Press, 1947.

WOODROW, H. The common factors in fifty-two mental tests. *Psychometrika*, 1939, 4, 99–108.

(Received October 31, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 12–18

A STUDY OF THE EFFECTS OF VERBALIZATION ON PROBLEM SOLVING[1]

ROBERT M. GAGNÉ AND ERNEST C. SMITH, JR.
Princeton University

Scattered throughout the literature on problem solving are occasional studies which are interpreted as indicating that acts of verbalizing during problem solving result in lessened problem solving effectiveness. While most investigators of such behavior may be inclined to expect a facilitating effect of "transfer of principles" to a final performance, studies have sometimes cast doubt upon the generality of such a finding, particularly in those instances where the principles are stated verbally.

Katona (1940), for example, found that a method which involved teaching verbal principles in solving matchstick problems to be less effective than a method of teaching by example. In a more recent study, Haslerud and Meyers (1958) found an experimental treatment in which verbally-stated principles of solution of cryptograms were given to *S*s to be less effective for solution of new cryptograms than was a treatment in which *S*s were required to discover solutions for themselves. Other findings supportive of this sort of conclusion are cited by Haslerud and Meyers (Hendrix, 1947) which suggest the superiority for transfer of "not-verbalizing" vs. "verbalizing" by *S*s themselves, in solving mathematical problems. In another study using matchstick problems, Corman (1957) failed to find significant differences in performance among groups given various amounts and kinds of verbal instructions.

Results like these contrast markedly with those of an older study by Ewert and Lambert (1932), which used as a problem the task of transferring discs graduated in size from one circle to another in a triangular configuration of three circles. In successive tasks, different numbers of discs (from three through eight) are placed in Circle No. 1, arranged in order of size with the largest at the bottom. The problem is to transfer the discs from Circle 1 to Circle 2 in the least possible number of moves, moving one at a time. One of the experimental groups in this study was given only instructions about the rules of the game; a second group was encouraged to try to find a general principle for solution; a third was given a verbally-stated principle of solution; while a fourth was given a principle plus a demonstration of the correct method with three discs. A large difference was found in the performance of groups who were given a verbally-stated principle of solution and the performance of groups who were not.

The three-circle problem is fairly difficult, and one can spend much time and many moves before discovering a principle which is truly general. Perhaps the reason for the striking contrast between the results of Ewert and Lambert (1932) and those of more modern investigators resides simply in this fact. To a participant in the problem, it seems natural indeed that his performance should improve once he knows the principle—improve not only on the problem he has attempted, but on others of a similar type. It will perhaps take more than one additional study to explicate these contrasting results.

It was our intention in the present experiment to make a further exploration into the effects of verbalizing on problem solving performance. We were interested particularly in the kind of verbalizing done by *S* himself during attempts to solve the problem, rather than by *E*.

Presumably, verbal principles provided in instructions must be repeated (perhaps to himself) by *S*, if

---

[1] This study was supported in part by funds granted by the Carnegie Corporation of New York. The opinions expressed are those of the authors, and do not necessarily reflect the views of that Corporation.

Data of the experiment were collected by the junior author under the direction and monitoring of the senior author. Portions of the data and results were described by Ernest Smith in his senior thesis.

they are to be effective. If we let *S* discover his own principles, in his own words, but require that he verbalize them, will this facilitate or interfere with problem solving? Marks (1951), for example, found no significant effects of providing *S*s with a typed list of principles ("elements of the problem"), but a high correlation (.83) between performance and vocalization by *S*s during solution. We were also interested in seeing whether we could establish the differences in performance suggested (but not confirmed) by Ewert and Lambert's (1932) results, of the effects of instructions to find and formulate verbally a general principle.

We chose to investigate these questions by measuring performance on a standard series of three-circle tasks of the sort employed by Ewert and Lambert, transfer to a final six-disc task of this type, and the adequacy with which *S*s could make verbal formulations of general principles. Specifically, the experiment compared the performance of groups of *S*s who solved two-, three-, four-, and five-disc problems successively, under four conditions representing combinations of two treatment variables: (*a*) a requirement to state verbally a reason for *each move* at the time it was made; and (*b*) instructions to search for a general principle which could be stated verbally after the tasks were solved.

METHOD

Materials.—The three-circle problem described by Ewert and Lambert (1932) was represented by three circles of 5 in. in diameter, drawn on a piece of stiff white paper, with their centers at the apexes of an equilateral triangle of side 7 in., and labeled A, B, and C. The discs were made of $\frac{3}{32}$ in.-aluminum, numbered 1 through 6, and graduated in diameter from $\frac{3}{4}$ in. to 2 in. A set of discs (two to six) is placed in Circle A, graduated with the largest at the bottom, and the problem is to move them all to Circle B so that they will be in the same order, in the smallest possible number of moves. Only one disc at a time may be moved, and it is never permitted to put a larger diameter disc on top of one with a smaller diameter. Under these rules, the fewest number of moves required (with any number n of discs) is $2^n - 1$.

The kind of principle which reduces this problem to a routine is the following: "If the number of discs is odd, move first to the circle to which you want to go eventually; if even, move first away from this circle. Continue by moving discs with odd numbers always in a clockwise direction, and the discs with even numbers always in a counterclockwise direction." There are, however, a number of other ways of formulating the second part of this principle, which are equally effective, although requiring more words.

Subjects.—The *S*s were 28 boys in Grades 9 and 10, who were assigned randomly to four experimental groups. Their ages were 14–15 yr., and their IQs were all above 110. They had volunteered to participate in studies of learning, but were paid an amount equivalent to prevailing rates of odd-job work. Each *S* was questioned closely to determine that he had no previous acquaintance with the problem, and was not used in the experiment if he had.

Procedure.—First, each *S* was shown the materials and given instructions about the rules of the game. The three-disc and the four-disc problem were then administered in succession, and each was carried to final solution (i.e., getting all discs in Circle B in the proper order). The *S* was told in each case what the minimal number of moves was. If he decided he had made a wrong move, he was permitted to go back to an earlier point in the solution, or to the beginning. The *E* made a count of all moves. The purpose of this exercise was to give all *S*s equal acquaintance with the problem, and also to provide data on equivalence of the groups.

Following this initial test, each *S* was assigned randomly to one of four conditions, each containing 7 *S*s, as follows:

Group V-SS (Verbalizing, Solution Set) was instructed to state aloud why they were making each individual move at the time they made it. In addition, these *S*s were instructed to try to think of a general rule by means of which they could tell someone how to solve these problems, which was to be solicited afterwards by *E*. Group V (Verbalizing, No Solution Set) was required to verbalize a reason for each move, but was not instructed to try to formulate a general rule for solution. Group SS (No Verbalizing, Solution Set) was not required to verbalize, but was instructed to try to formulate a rule. Group No (No Verbalizing, No Solution Set) was simply told of the problem to be presented and its ground rules, with no additional instructions.

With these instructions, *S*s were kept at the task until they achieved a single final solution for two, three, four, and five discs in succession. One-minute rest periods were interposed between successive tasks. Number of moves was counted in each case. Of course, the making of moves was slower for those *S*s who were required to verbalize, usually to the point of slight annoyance. Following this learning session, after approximately 3 min. rest, all *S*s were presented the six-disc task as a final test. They were told that their time to solution would be measured (with a stop watch), as well as the number of moves taken. No verbalizing was required.

Recording.—For the initial test (using three and four discs), *E* simply kept a record of the number of moves made by each *S*.

During administration of the main experimental treatments (two, three, four, and five discs) number of moves was recorded. For those groups instructed to verbalize each move, *E* recorded on a prepared record sheet a brief phrase indicating the verbal statement made by each *S*. No attempt was made to make exact verbatim reproductions of what *S* said. Whenever an unusual reason was given, however, *E* attempted to record its meaning fully.

On the final test, *E* recorded the number of moves and the time required to achieve solution. After the final test was concluded, each *S* was asked to give a rule or rules for doing the problem with any number of discs, as if he were telling it to a person ignorant of the solution. The *E* recorded these verbally stated principles.

RESULTS

Initial Test

The means and *SD*s for number of moves made by each of the four groups on the initial test composed of tasks with three and four discs were as follows: V-SS: $M = 26.4$, $SD = 4.5$; V: $M = 28.6$, $SD = 3.7$; SS: $M = 26.6$, $SD = 3.6$; No: $M = 27.9$, $SD = 4.9$. These results appeared to insure an acceptable degree of comparability ($F = .70$, $df = 3/24$; $P > .05$).

Practice Session

Mean performance curves for the four groups of the experiment are shown in Fig. 1. These depict mean number of moves in excess of the minimum for the problems with two, three, four, and five discs administered in sequence. The same measure on the final test is also shown for each group as a terminal point.

It may readily be seen that a difference among the groups, associated with the Verbalization variable, began to appear as early as the three-disc practice, and showed itself as an ever-widening difference thereafter. On the other hand, no differences of consequence appear

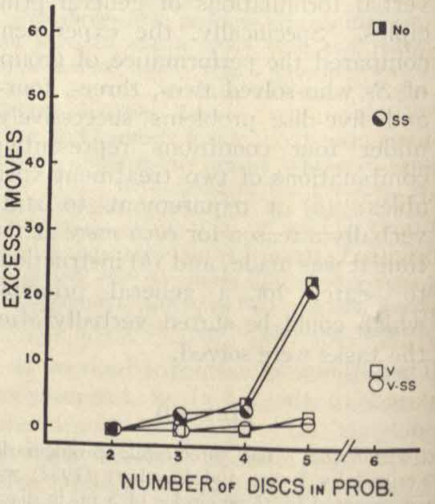


FIG. 1. Performance curves showing moves in excess of minimum required for successively administered problems of two, three, four, and five discs for the four groups of the experiment. (Values are shown for the final six-disc problem as a terminal point.)

in the performances of groups differentially treated with respect to a Solution Set, i.e., being told to look for a principle to be stated verbally. These apparent trends of the data were confirmed by an analysis of variance of the data for Trial 5. Since heterogeneity of variance was found in the raw data, they were transformed to logarithms. In this state, the null hypothesis regarding differences in variances could be rejected. An analysis of variance on these transformed data indicated significant treatment effects ($F=6.67$, $df=3/18$; $P < .01$). When t tests were applied to the pairs of differences between means these were found to be significant at less than the .01 level for each of the verbalization groups vs. each of the nonverbalization groups.

In contrast, neither of the pairs of groups differentiated by the presence or absence of the SS instructions yield means which are significantly different from each other. Verbalization made a significant difference, Solution Set instructions made none.

Verbalization during practice.—The verbalization done by *S*s who were instructed to give a reason for every move tended to fall into certain standard categories, with rare exceptions. There were those which were oriented toward *single* moves, like (*a*) "only possible move;" (*b*) "just to try it;" or (*c*) "don't know" (this one was quite infrequent). Then there were those which anticipated to the extent of *two* moves, such as (*a*) "to get at the larger disc;" (*b*) "to free up a space." There were also instances of reasons which apparently anticipated *sequences* of moves. One of the relatively frequent ones was (*a*) "move as with a three-disc sequence;" also in this category should perhaps come (*b*) "if disc is odd-numbered, move to circle B," and (*c*) "if disc is even-numbered, move to circle C." Finally there were some expressions of truly *general* principles, such as (*a*) "move odd-numbered discs in the clockwise direction," and (*b*) "move even-numbered discs in the counterclockwise direction."

On the whole, the content of these "reasons" was by no means startling. It was not particularly surprising, considering the results of earlier investigators of thinking, that this form of thinking out loud was not very revealing about the nature of internal processes. A close examination of these data did not enable us to invent a method of relating specific types of verbal response to specific stages of the problem, except for the fact that the more general principles tended to occur later in practice than did the less general ones.

Final Task Performance

Mean number of moves in excess of minimum, and mean times (min.) to solution, for each of the four groups of the experiment on the final six-disc task are shown in Table 1. Here it can be seen that under standard conditions for all the groups, the differences in moves which appeared on the five-disc task are even greater on the six-disc task. Again the variances are markedly different between the verbalization and nonverbalization pairs of groups. A logarithmic transformation was found to make possible the assumption of homogeneity of variance. An analysis of variance performed on these transformed data confirmed significant treatment effects ($F=9.13$, $df=3/18$; $P < .01$). Application of t tests to differences between individual means indicated significance at the .01 level for comparisons of the verbalization and nonverbalization groups. Dif-

TABLE 1

Means and *SD*s for Number of Moves in Excess of Minimum, and for Time of Performance, on the Final Six-Disc Task for the Four Groups

Measure	Groups			
	V-SS ($N=7$)	V ($N=7$)	SS ($N=7$)	No ($N=7$)
Moves in excess of minimum				
Mean	7.9	9.3	48.1	61.7
SD	8.6	7.9	34.8	40.2
Time (min.)				
Mean	4.2	3.8	10.1	10.0
SD	2.2	2.0	4.1	5.5

ferences between means for Groups SS and No, and Groups V-SS and V, are not significant.

An entirely similar set of comparisons exists with the means of time to solve, shown in Table 1. Here the F test for the total data is significant at lower than the .05 level ($F = 4.90$, $df = 3/18$). The t test indicates significance of differences of means associated with the contrast of verbalization vs. no verbalization (V-SS vs. SS; V-SS vs. No; V vs. SS; V vs. No) at better than the .05 level. Other differences in means are not significant.

Stating the Verbal Principle

Records of statements made by *S*s in response to the instruction to state verbally how to solve these problems were examined by the 2 *E*s independently, and rated "Inadequate," "Partial," or "Complete." The following agreements were made beforehand: "Inadequate" was to mean any principle which was incorrect, irrelevant, or so fragmentary that it would be of no aid to anyone in finding solutions. "Partial" meant one of the two major parts of the principle, concerned with starting (move odd-numbered disc to the goal circle, etc.) or with procedure during solution (even numbers move clockwise, etc.) "Complete" meant both of these major parts.

With these criteria, it was quite easy to classify the principles the *S*s stated, and there were no differences in judgment of the 2 *E*s. These results are shown in Table 2.

Considering first the dichotomy Inadequate vs. Partial or Complete, the Fisher exact probability test was applied to comparisons of the experimental treatments. The probability of the observed contrasts between V-SS and No, and between V and No were found to be .002; between SS and No, .014. All other comparisons yielded high probability values. So far as being able to state even *partial* verbal principles, Group No is significantly poorer than those who verbalized; for those who did not verbalize, there is a significant effect of instructions to look for a general principle. Second, the dichotomy Inadequate or Partial vs. Complete was considered by this same method. Here the only low probabilities were .036 between Groups V-SS and SS and also Groups V-SS and No. The group required to verbalize as well as to try to formulate a verbal principle,

TABLE 2

Number of Inadequate, Partial, and Complete Verbal Principles Stated by *S*s in Each Group

Group	N	Number of Instances of Principles Judged as:		
		Inadequate	Partial	Complete
V-SS	7	0	3	4
V	7	0	5	2
SS	7	1	6	0
No	7	6	1	0

was thus found to be superior to Groups SS and No in the formulation of *complete* verbal principles. It would appear that verbalizing during practice is the most important factor accounting for this difference.

DISCUSSION

Our results indicate that requiring individuals to verbalize while practicing the three-circle problem is a condition which is significantly related to superior performance in problem solving, which begins to show itself during practice with two-, three-, four-, and five-disc tasks, and is maintained on a final six-disc task. There is a strong suggestion also, that verbalizing is the most important factor at work in producing a greater number of individuals who can, at the close of practice, state fully adequate verbal principles of task solution.

Instructions to try to formulate verbal principles for solution of these tasks appear to have no effect on performance of the task, and also an insignificant effect (by themselves) in producing a greater frequency of fully adequate principles. They do, however, produce more verbal principles which are at least partially adequate, than is the case when they are omitted. It will be apparent in interpreting these results that we cannot be certain how great a contrast in "set to look for solution" was actually achieved by the instructions used. After all, the noninstructed groups as well as the instructed ones underwent an initial period of testing on the three- and four-disc tasks, and may have generated in themselves a set to look for a general solution.

Surely the most striking finding is the effect of verbalization by the *S*s during practice. Such verbalization is of course very different from that which may result from providing *S*s with verbally stated principles of solution, as was done in the studies of Ewert and Lambert (1932) and Haslerud and Meyers (1958). Whatever principles were used by the individuals in the present experiment, they must have been discovered by the *S*s themselves. In this respect, at least, the present findings are not inconsistent with those of Haslerud and Meyers, nor with other results indicating the effectiveness of self-discovered principles for problem solving (e.g., Gagné & Brown, 1961).

It is perhaps of some importance to emphasize that the individuals who were required to verbalize also took more time to make successive moves. We cannot tell from these data, of course, whether this more deliberate pacing of the task may have had some effects on the performance. The additional time between moves, however, was "filled" time, taken up entirely with the act of verbalization. Thus it would not be reasonable to suppose that the verbalization groups had any greater opportunity than did the nonverbalization groups for the deliberate rehearsal and recall of successful and unsuccessful moves. Verbalization is the most obvious variable at work; if other events are contributing to the results because of their association with time differences, it is not immediately apparent what they are.

As we pointed out, the content of the verbalizing during practice was fairly pedestrian and to some extent routine, so that it could be readily categorized. What then accounts for its effect on problem solving? In answering this question, we have no theory to call upon. It would appear that requiring verbalization somehow "forced the *S*s to think." In other words, this treatment may have had the effect of constantly prodding the *S*s to think of new reasons for their moves, particularly since they may have gotten a little tired of giving "the same old reasons" over again. This conception of the treatment would assign it a role similar to that of instructions to think of new word associates, as in the studies of Maltzman, Simon, Raskin, and Licht (1960). But this is a speculation which obviously does not come directly from present data. It would, however, be an interesting route for follow-up studies to take; namely, testing whether instructions to "think of new reasons," to be stated verbally, would have a facilitative effect on the solution of a task like that used here.

SUMMARY

A study was conducted to determine the effects of (*a*) requiring *S*s to verbalize during practice, and (*b*) instructions to find a general principle to be stated verbally, on problem solving performance. The three-circle task described by Ewert and Lambert (1932) was used. The *S*s were 28 ninth and tenth grade boys, divided randomly into four groups of 7 each. Following the initial test on tasks with three and four discs, *S*s were assigned to four groups given different treatments representing combinations of the two experimental variables. The practice session consisted of two-, three-, four-, and five-disc tasks administered successively. Records were kept of the verbalizing, for those groups who engaged in it, and of number of moves to achieve solution, for all groups. Following this, a final task using six discs was administered to all *S*s under standard conditions, and a record was made of number of moves as well as time to solution.

Significant differences were found between the scores of those groups who were required to verbalize and those who were not, in terms of number of moves. Similar differences based on time scores were found in the contrast between verbalization and nonverbalization groups. Differences between other pairs of groups were not significant.

Comparisons of success in stating verbal principles applicable to the solution of three-circle problems were made among judged categories of Inadequate, Partial, and Complete, for all groups. A significant difference was obtained between verbalizing and nonverbalizing groups, and one which favored the "instructed" groups when the dichotomy Inadequate or Partial vs. Complete was tested.

The results appear to indicate that requiring *S*s to verbalize during practice has the effect of making them think of new reasons for their moves, and thus facilitates both the discovery of general principles and their employment in solving successive problems.

REFERENCES

CORMAN, B. R. The effect of varying amounts and kinds of information as guidance in problem solving. *Psychol. Monogr.*, 1957, **71**(2, Whole No. 431).

EWERT, P. H., & LAMBERT, J. F. Part II: The effect of verbal instructions upon the formation of a concept. *J. gen. Psychol.*, 1932, **6**, 400–413.

GAGNÉ, R. M., & BROWN, L. T. Some factors in the programing of conceptual learning. *J. exp. Psychol.*, 1961, **62**, 313–321.

HASLERUD, G. M., & MEYERS, S. The transfer value of given and individually derived principles. *J. educ. Psychol.*, 1958, **49**, 293–298.

HENDRIX, G. A new clue to transfer of training. *Elem. sch. J.*, 1947, **48**, 197–208.

KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

MALTZMAN, I., SIMON, S., RASKIN, D., & LICHT, L. Experimental studies in the training of originality. *Psychol. Monogr.*, 1960, **74**(6, Whole No. 493).

MARKS, M. R. Problem solving as a function of the situation. *J. exp. Psychol.*, 1951, **41**, 74–80.

(Early publication received July 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 19–22

COUNTERTRAINING OF A SIMPLE SKILL WITH IMMEDIATE AND 1-WEEK DELAYS OF INFORMATIVE FEEDBACK [1]

FRANCIS J. RYAN [2] AND EDWARD A. BILODEAU

Tulane University

Delaying knowledge of results (KR) or informative feedback by 7 days has the effect of interpolating a 7-day layover between successive responses. Such spacing retards the learning of a simple, discrete motor response. The effect of delay of KR is another question, as was made clear in a series of five experiments by Bilodeau and Bilodeau (1958). They found no effects attributable to the time between R and KR, and a negative effect attributable to the time between one R and the next. The interresponse intervals and delays of KR ranged from 20 sec. to 7 days. The interval between KR and the next R (post-KR interval) also ranged up to 7 days.

Their Exp. 4 and 5 are particularly relevant here. Experiment 4 used a 1-day interval between Rs, and showed a group with KR delayed by 1 day to be superior to a group with no delay. The most plausible hypothesis was that delay within a relatively massed series of Rs was unimportant while to delay KR would be best with long interresponse intervals, i.e., delay of KR interacts with interresponse interval. Experiment 5 was designed for confirmation. Two groups employed a 7-day interval between Rs, one was given KR immediately; the other was delayed for 7 days. The results were disappointing, though not unexpected. Neither group learned more than a trivial amount, so there was no opportunity for a difference as a function of delay. The wide spacing of Rs had again shown its powerful effect. Experiment 5 had too few *S*s and trials. Both were in short supply since R was scheduled at the rate of one per week. The present experiment used a greater number of *S*s and Rs than did Exp. 5.

Still another consequence of the week-long interval was the large between-*S* variance; that is, a great deal of hunting behavior was shown. The problem of the unwieldy variances led to a review of techniques for regulating learning in human beings. No existing task for studying the learning process in human *S*s was judged suitable. For example, too often, conditioning takes place too fast or not at all. Control over extinction rates poses even greater problems. Further, in the usual task, control over the variances is left to luck.

Some control over between-*S* variance is possible. The present solution is to train *S*s to homogeneous performance. After *S*s are performing similarly, the experimental treatments are introduced at the same time that the task is changed to require new Rs instead of the previously trained R.

The writers wish to introduce a practical countertraining technique to the motor skills area and propose that informative feedback is the natural vehicle toward this end. For the training period, *S*'s feedback error will be computed with regard to a particular target. After training is completed, the target is shifted systematically, the only indication of change lying in KR. To keep KR errors to a minimum, *S* must learn to modify successive Rs. The experimental variable—delay of KR—is introduced simultaneously with the target changes.

---

[1] This work was supported by a grant from the National Science Foundation to E. A. Bilodeau, principal investigator. It also represents an abridged version of a thesis submitted to the Graduate School of Tulane University in partial fulfillment of the requirements for the Master's Degree.

[2] Now at the State University of Iowa.

METHOD

Subjects.—Seventy students at Tulane University volunteered for eight sessions at weekly intervals. They were assigned to two groups and run individually according to a counterbalanced schedule. To assure better attendance, *S*s completing the 8 weeks were promised $3.00; the 6 best performers expected prizes ranging from $10.00 to $50.00. Nine *S*s failed to keep appointments. Also voided was an injured *S*, 1 misrun by *E*, and 2 who failed to meet a pre-established training criterion.

Apparatus.—Manual Lever C, a modification of an earlier device developed by Bilodeau and Ferguson (1953) was used. It is a stick or lever mounted inside a 23 × 17 × 17 in. wooden cabinet. Lever displacement (*S*'s response) is measured by means of a scale on the back of the cabinet. The *S* was unable to see the lever and scale.

The task was to move the lever from a zero (starting) position to a second position (target). The *S* attempted to hit the target on each move. A response 1° short of the target was reported as "short by 5 units," a move 4° beyond the target was reported as "long by 20 units," etc. Units were not defined for *S*.

Procedure.—On Day 1 (training), 8 Rs were required. The first 7 of these were all followed by immediate KR. The experimental variable was introduced after R8 when half of the *S*s were switched to a 7-day delay schedule and the remaining *S*s continued with immediate KR. R9 was required on Week 2, R10 on Week 3, and so on through Week 8. For purposes of reference: the first 8 Rs have been assigned to training; the countertraining procedures begin immediately after the termination of R8, and the countertraining effects begin (potentially) with R9 (Week 2).

Three substantial changes in procedure followed R8: (*a*) the interresponse interval was lengthened to 7 days; (*b*) the temporal locus of KR was changed, a week being interpolated between R and KR for Group D, and a week being interpolated between KR and R for Group I; and (*c*) the target position was decreased by 4° per trial. KR was delayed by a trifle (2 or 3 sec.) for Group I; for Group D, the post-KR period was held close to 15 sec. These temporal intervals minimize the confounding problem (Bilodeau & Bilodeau, 1958), since the KR delay is a small fraction of the intertrial interval for Group I, and the 15-sec. post-KR interval is also a small fraction of the intertrial interval for Group D. Elaborate precautions were taken to withhold any cue of the target changes other than that provided by KR.

Each *S* sat facing the lever for about 15 sec. during a demonstration of standard posture, grip, and rate of movement. Next, *S* was coached to avoid moving too fast (reduces kinesthetic cues) and moving too slow (produces chatter). The task was described and *E* stated that KR would assist in learning it. A "practice" move of "about half an inch" was then called for (R1) and immediate KR was given. R2 through R8 followed at 20-sec. intervals.

Following R8, KR was withheld from Group D, both groups waited 15 sec. (as if for another R) and were dismissed. The next week, *E* reminded *S* of the task, posture, rate of movement, and then gave KR to Group D. All *S*s waited 15 sec. before making R9. After responding, *S*s of Group I were given immediate KR and all *S*s waited 15 sec. before dismissal.[3]

All *S*s were taught to make a movement from 26° below the vertical to a position 10° beyond the vertical, an arc of 6.9 in. Using the preferred hand, half the *S*s moved from left to right, and half from right to left. If a delayed *S* asked for KR after responding, or if an immediate *S* sought a reminder of KR just before executing a response (such questions were rare), he was told that *E* lacked time to figure out the score right now, or that *E* had forgotten.

RESULTS

Figure 1 shows the mean amplitude of R for all 15 trials. The corresponding *SD*s have been placed in Table 1. Both Groups I and D learned the training task fairly quickly as is shown by the increasing means and decreasing *SD*s. The eight trials of training were just about sufficient to produce maximum accuracy (means) and near minimum variability (*SD*s). Further, Groups I and D, having been trained alike, behaved quite alike.

---

[3] After R15 of Week 8, both groups were given immediate KR, held 15 sec., and dismissed. One hundred and sixty days later the *S*s were brought back for a test of recall. The recall-test data were not part of the thesis work and are not reported here.

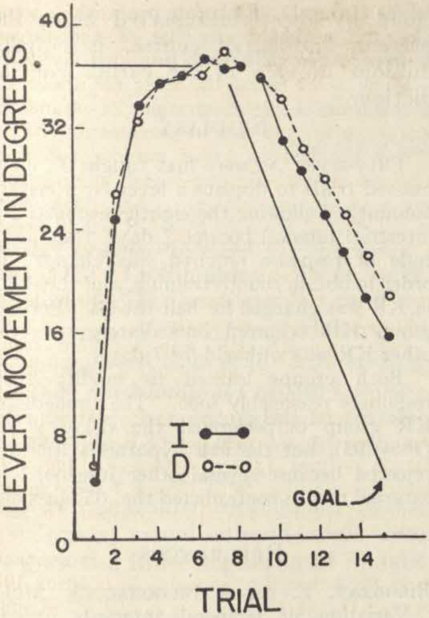


FIG. 1. Mean response amplitude for training (Trials 1–8) and for countertraining (Trials 9–15) for Groups I and D.

The means of the countertraining trials (9–15) show both groups learning to reduce the amplitude of R with successive trials; the mean square for trials was significant ($P < .01$). Though R was at the rate of one per week, there can be no doubt of a powerful learning effect in the means of both groups.

The solid line in Fig. 1 represents the loci of the targets. Comparing the location of Curves I and D to that of the target, we see that both groups lagged behind with the greater lag shown by Group D ($F = 5.01$; $df = 1/55$; $P = .05$). There is also a fairly constant difference between I and D; the interaction term produced an F less than 1.0. Group I has less than a one-trial advantage over D. This advantage first appears on Trial 2 of countertraining and does not increase thereafter. The differences between the means of Groups D and I (D − I) on Trials 9–15 were −0.2, 3.7, 1.4, 2.7, 2.7, 3.0, and 3.4. The overall difference is 2.4.

TABLE 1

*SD*s OF RESPONSE AMPLITUDES FOR GROUPS I AND D

Trial	I	D	Trial	I	D
1	2.1	3.9	9	7.9	6.4
2	18.2	14.5	10	6.9	5.3
3	12.5	10.7	11	5.1	4.5
4	9.2	8.0	12	5.1	5.2
5	5.6	6.1	13	5.2	5.5
6	8.6	5.1	14	6.0	6.0
7	3.4	3.6	15	8.1	10.9
8	3.3	2.9			

LIBRARY
1.9.70
256

Examination of the *SD*s in Table 1 shows similar variabilities from I to D. The major effect occurs between Trials 8 and 9 in that both groups become more variable as the intertrial interval was lengthened from 20 sec. to 7 days. There is no special effect at Trial 10 as there was with the means. During the last 2 weeks a few *S*s began to perform erratically.

The mean absolute errors were also examined carefully, but they are not reproduced here because the findings were highly similar to those already reported. Suffice it to say that the target moved 4° per week whereas the *S*s modified R by about 2° per week, the Group I *S*s showing a very small advantage on Trial 10.

Comments from *S* were not solicited. However, 6 of the 27 *S*s in Group D and 4 of the 30 *S*s in Group I made some remark about the moving goal. As in previous studies (Bilodeau & Ryan, 1960), almost no one complained about, or remarked on, the delay of KR.

DISCUSSION

This experiment was designed to provide further evidence on an earlier suggestion that the damage done by very long intertrial intervals may be lessened by delaying KR. No support for this idea was found.

The difference between Groups I and D at the .05 level and the insignificant interaction of I and D over trials were the two major findings. In five earlier studies there was no significant superiority of I over D, and one of D over I. Significance beyond the .01 level should be and was required for rejecting the null hypothesis; hence the present .05 difference is ascribed to error and chance. In addition to the previous studies, there were also internal reasons for the .01 criterion: (*a*) If delay of KR hinders the learning of this skill, the difference between Curves I and D should increase with trials. There was no evidence for such an interaction. (*b*) The difference between I and D did not originate with the first week of delay. Those who choose to reject the null hypothesis must propose a delayed effect. (*c*) Statements volunteered by *S*s contained very few references to KR—its delay and changing nature.

Bilodeau and Bilodeau (1958) have reported that a group with a 1-day intertrial interval and a 1-day delay of KR was superior to the group with no delay and a 1-day intertrial interval. The present experiment is similar, but the result is not. The treatment differs in that the intertrial interval is longer and the task superficially different. The previous finding has been placed under a cloud by the data reported here; the present finding, of course, is equally dubious in view of the earlier contradiction.

SUMMARY

Fifty-seven *S*s were first taught in eight massed trials to displace a lever by a certain amount. Following the eighth response, the intertrial interval became 7 days, the amplitude of response required was changed in order to obtain countertraining, and the delay in KR was changed for half the *S*s. For one group KR occurred immediately; for the other KR was withheld for 7 days.

Both groups learned to modify their responses reasonably soon. The immediate-KR group outperformed the delay group ($P = .05$), but the null hypothesis was not rejected because several other internal and external results contradicted the .05 outcome.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

BILODEAU, E. A., & FERGUSON, T. G. A device for presenting knowledge of results as a variable function of the magnitude of the response. *Amer. J. Psychol.*, 1953, **66**, 483–487.

BILODEAU, E. A., & RYAN, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *J. exp. Psychol.*, 1960, **59**, 414–419.

(Received September 1, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 23–28

A TEST OF THE RESPONSE PROBABILITY THEORY OF PERCEPTUAL DEFENSE[1]

MICHAEL J. GOLDSTEIN
University of California, Los Angeles

Two recent reviews on the influence of motivation upon perception (Eriksen, 1958; Goldiamond, 1958) have concluded that the dilemma posed by phenomena of perceptual defense is more apparent than real. The finding of lower thresholds for anxiety-arousing words (when compared to thresholds for neutral words) would appear to imply the existence of a supersensitive scanning mechanism which first scans visual stimuli and then inhibits the full recognition of images likely to arouse anxiety. Eriksen (1958) and Goldiamond (1958) argue that another, more parsimonious explanation can account for the phenomena of perceptual defense, an explanation which emphasizes the operation of response variables. These authors suggest that as a consequence of prior conditionings, *S* develops differential probabilities of using various words as responses in a perceptual situation. In the case of perceptual defense, responses which have become conditioned stimuli for anxiety will be less likely to be evoked under conditions of marginal perceptual stimulation than other stimuli of similar physical contour which have not shared this association with noxious stimulation. This lowered probability of evocation (negative response bias) of anxiety-arousing words can account for the perceptual defense effect.

If *S* is going to call a particular class of words less frequently than another, then the possibility of a chance correspondence between his response and the correct stimulus according to *E*'s answer sheet is less likely for the undercalled class than for the overcalled class of words. The higher threshold for anxiety-arousing words thus becomes a simple artifact of the operation of response factors and does not require the postulation of a prescanning mechanism in the visual system.

This response probability theory emphasizes the necessity of separating two sources of variance in perceptual recognition scores, one component which appears to be independent of the presence of the perceptual stimulus and a second source which appears to require the presence of the perceptual stimulus before certain effects can be noted. Goldiamond (1958) refers to these two components as extraneous variance and perceptual variance and argues for the necessity of isolating them. The purpose of the present report is to attempt an experimental analysis of the phenomena of perceptual defense in an effort to isolate these two sources of variance. In attempting this analysis, four basic questions are asked: (*a*) Does a negative response bias exist for anxiety-arousing words when no discriminative stimulus is presented? (*b*) If present, is such a bias of sufficient magnitude to account for the differences in accuracy, when no discriminative stimulus is presented? (*c*) What changes in response bias occur when a discriminative stimulus is present, and are they sufficient to account for all of the variance in accuracy scores? (*d*) What effect does increasing the amount of available information have upon response bias and accuracy?

---

[1] This study was supported by a grant from the Faculty Research Committee, University of California, Los Angeles. An abridged version of this paper was presented at the Western Psychological Association meetings, April 1959 in San Diego, California. The author would like to express his appreciation to Samuel Himmelfarb for running *S*s and analyzing data and to Norman Anderson for his generous help on statistical matters.

METHOD

Subjects.—There were 60 college student *S*s, 30 of each sex. Each *S* was assigned at random to one of three experimental groups: a Stimulus Absent (SA) group, a Stimulus Present (SP) group, and an Increasing Information (II) group, with the restriction that there were 20 *S*s in each group.

Selection of anxiety-arousing and neutral stimuli.—In a preliminary session each *S* was administered a 117 item word association test. The stimuli in this test were constructed in triads, each word designed to be potentially anxiety arousing being paired with two other words of the same frequency of usage, same length, same starting letter, and neutral in content. The triads were not adjacent in the test, but were randomly interspersed throughout. The *S* was instructed to respond to the stimulus word with the first word that came to mind; his reaction time (RT) (in .1 sec.) and his verbal response were recorded. After completion of the test, the list was read a second time and *S* was instructed to give the same response that he gave on the first administration.

A word was selected as anxiety arousing for *S* if there was a prolonged RT to the word and a reproduction fault on the second administration. The RT for the anxiety word had to be several seconds longer than the RT to the two other members of its triad for a word to be considered anxiety arousing. For each anxiety-arousing word selected, a neutral word was selected from the same triad for which no block or reproduction fault was evident. Four anxiety-arousing and four neutral words were selected for each *S*. Since both classes were selected from the same triad, the final list of eight words involved four pairs of words with identical frequencies of usage, length, and starting letter.

Perceptual task.—Following the word association test, *S* was scheduled to return within 48 hr. Upon his return *S* was introduced to the Gerbrands tachistoscope and an estimate of his visual sensitivity was obtained. The *S* was shown a list of eight words and was asked to learn them. The words were all of extremely high frequency of usage and included three key words: LIGHT, DOLLAR, and RETURN. Although *S* was told that all eight words were to be flashed in the tachistoscope, only the three key words were flashed according to a prearranged random order. These key words were each presented twice at a series of increasing exposure times until a setting was reached at which *S* could identify any one of the three words on two nonconsecutive trials. This exposure time was considered a lower bound of the range of *S*'s visual sensitivity. Once this exposure time was determined, the tachistoscope was then set at 50–100 msec. below this figure.

For Groups SA and SP, the exposure time set for *S* was used for all trials of the experiment; however, Group II received the first block of 32 trials at this setting and each subsequent block of 32 exposures at a setting 50 msec. above the previous block's level.

Following the determination of sensitivity *S* was shown another list of eight words which had been selected from his word association performance and was told that now only these eight words would be flashed in the tachistoscope. A minute was given to *S* to look over the list, after which it was attached to the tachistoscope. Specifically, *S* was instructed:

> This is a study dealing with perception. Here is a list of eight words which will be presented one at a time, in a random order, in this machine. You are to look into the eyepiece and judge which of the eight stimuli is present on each trial by referring to your list. The words will be presented rapidly so that at times it will be very difficult for you to see the words and you will frequently have the feeling that you are guessing. Don't be disturbed by this, just guess on every trial.

The *S*s in Group SA were instructed that words were to appear in the tachistoscope, but *E* actually flashed a single nonsense figure (five typewritten hashmarks) for the 128 trials of the experiment. Group SP actually received a word on each of the 128 trials according to a prearranged random order. The only limitation on the random order was that both classes of words (anxiety arousing and neutral) had to be presented equally often and that in each of the four blocks of 32 trials, each of the eight words had to appear equally often. These limitations were placed on the random order in order to provide for an analysis of a possible effect over trials upon bias and accuracy. For Group II words were flashed on each of the 128 trials just as for Group SP.

The experiment was conducted in a windowless, black, experimental room, the only illumination coming from a 7-w. bulb.

Pseudoaccuracy scores.—To answer the question of whether response bias can produce the perceptual defense effect, it was necessary to compute pseudoaccuracy scores for Group SA. Pseudoaccuracy scores were obtained by scoring *S*'s responses as correct or incorrect according to the random order of stimulus presentations actually used with the other two experimental groups. These scores would be expected to vary about a chance mean of $12\frac{1}{2}\%$ correct.

RESULTS

The results were analyzed in terms of performance over all 128 trials and in terms of performance in successive blocks of 32 stimulus presentations. The data for over-all performance will be presented first, as they correspond to the kind of data typically reported in perceptual defense studies. It should be made clear, however, that the results are not described in terms of thresholds, but in terms of accuracy scores over successive runs at a single stimulus presentation or over blocks of increasing exposure times. This was done to avoid some of the pitfalls judged to be present in the classical ascending method of limits, in which termination of series by a correct response undoubtedly alters the system of response biases for the remaining words in the set.

First we turn our attention to the question of whether response bias exists when no discriminative stimulus is present. The data from Group SA were analyzed in terms of the total number of emotional words called by each *S* over the 128 trials. Since 64 calls would indicate no bias, each *S*'s bias score was subtracted from 64 to yield a difference score. The mean difference score for Group SA was -4.55 (-2.3 for men, -6.8 for women), indicating a negative response bias with regard to emotional words. Analysis of variance of these difference scores revealed a significant F (6.53 for 1/18 *df*, $P < .025$) for the mean, but no significant F for sex of *S*.

The next issue was whether this degree of response bias can bring about differences in accuracy scores which would ordinarily be interpreted to be due to perceptual factors. Each *S*'s record was scored for pseudoaccuracy (i.e., the number of times he would have identified the stimulus had it actually been presented). Separate scores were obtained for anxiety-arousing and neutral words. An *S*'s mean pseudoaccuracy score for neutral words was subtracted from his pseudoaccuracy score for anxiety-arousing words, yielding a distribution of difference scores. The mean difference was -2.45 ($-.3$ for men, -4.6 for women), suggesting lower pseudoaccuracy scores for the anxiety-arousing words. Analysis of variance of the difference scores revealed an F of 8.87 for the mean ($P < .01$ for 1/18 *df*), indicating that the mean difference is statistically different from zero, and an F of 6.81 for sex ($P < .025$ for 1/18 *df*). The latter finding indicates greater differences in pseudoaccuracy scores for women than for men. The analysis of pseudoaccuracy indicates clearly that the negative response bias is sufficient to produce significant differences in pseudoaccuracy. Had a real perceptual stimulus been presented, we would have erroneously concluded that the poorer rate of recognition for anxiety-arousing words (compared with neutral words) was a function of perceptual factors rather than response factors!

Effect of presenting a discriminative stimulus.—The fact that response bias can produce the results described above is no assurance that this variable is the sole one in operation when a discriminative stimulus is present. In an effort to study the possibility that other factors are operating, the data from Group SP were analyzed in terms of response bias and accuracy. Bias was computed in the manner described above. The mean difference score for Group SP was -4.90 (-5.30 for men, -4.50 for women), indicating a negative response bias, but no tendency for the bias to relate to sex of *S*. An analysis of variance of the bias difference scores revealed an *F* of 16.32 ($P < .001$, for 1/18 *df*) and an insignificant *F* for sex. This analysis indicates that there exists a significant tendency to undercall anxiety-arousing words when a discriminative stimulus is flashed in the tachistoscope.

A similar analysis was done on the accuracy scores for anxiety-arousing and neutral words. The number of stimuli in each class correctly identified by *S* over 128 trials was computed. A difference score was then computed for each *S* between his accuracy scores for anxiety-arousing words and his accuracy score for neutral words. The mean of these difference scores was -4.50 (-6.10 for men, -2.9 for women) which was statistically different from zero ($F = 12.62$, $P < .005$ for 1/18 *df*.) This merely duplicates the well known perceptual defense effect in which *S* shows poorer rates of decoding for anxiety-arousing words than for neutral words.

Comparison of stimulus Groups SA and SP.—A stated purpose of the present research was the isolation of perceptual variance from response variance. Since the data from Group SA represented a "pure" measure of differences in accuracy due to response bias, a test was made of whether the difference in accuracy found in Group SP was significantly greater than could be accounted for by response bias alone. Using a distribution of difference scores, an analysis of variance was done to determine if the mean difference in accuracy between anxiety words and neutral words in the Group SP (-4.50) was significantly greater than the mean difference in accuracy (-2.45) found in Group SA. An analysis of variance of these difference scores revealed that the variance between groups was not statistically significant ($F = 1.88$, $P > .20$ for 1/38 *df*). Thus, the magnitude of the perceptual defense effect found in the presence of a discriminative stimulus is not significantly greater than the magnitude that would be expected on the basis of response bias alone.

Performance over trials.—The order of presentation of stimuli was designed to permit a four-block analysis over trials, with 32 stimulus presentations in each block. This was done to examine any trends in both call rate and accuracy as a function of repeated presentations of anxiety-arousing and neutral stimuli. Figure 1 presents the data for response bias over blocks for the three experimental groups. It can be seen that the trends over blocks is different for the groups which received stimuli than for Group SA. Both Groups SP and II exhibit an increase of negative response bias over blocks, while the data for Group SA suggest a gradual decline in bias. A statistical comparison of response bias was done on the trends for Groups SA and SP in the manner suggested by Grant (1956) for repeated measurements

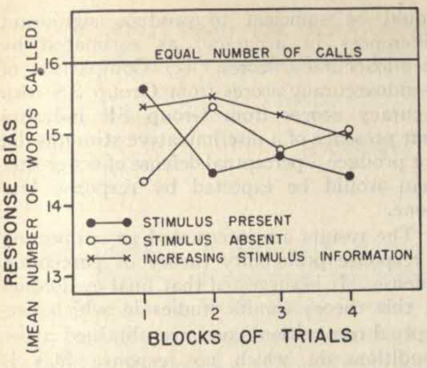


FIG. 1. Mean number of anxiety-arousing words called over the four blocks of the experiment by *S*s in the three groups.

data. The analysis was restricted to the linear component scores. The mean linear score for Group SA was +4.1 (+0.10 for men; +4.0 for women) and the mean linear score for Group SP was −6.5 (−5.0 for men; −1.5 for women) supporting the visual impression of the two curves. However, the analysis of variance of these linear scores is not significant for groups ($F = 1.170$, $P > .10$ for 1/36 *df*), sex ($F = 1.68$, $P > .10$) or the Groups × Sex interaction ($F < 1$), all tests for 1/36 *df*.

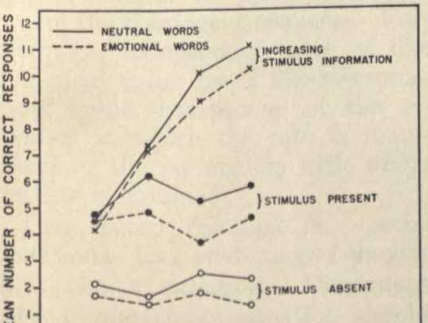


FIG. 2. Mean number of anxiety-arousing (emotional) words and neutral words correctly identified over the four blocks of the experiment by *S*s in the three groups.

A similar analysis by blocks was done on the accuracy scores and is presented in Fig. 2. It can be seen that very little increase in accuracy scores occurred with repeated presentations of the stimulus (Group SP) and that improvement in accuracy is directly related to longer exposure times of the stimulus (Group II). Further, differences in accuracy appear between anxiety-arousing and neutral words over blocks of trials, even when the later blocks of trials imply longer exposure times. Rather than extinguishing over the course of the experiment, the perceptual defense effect appears to become enhanced with practice! A comparison of the trends in Groups SA and SP was done in the manner described above for bias scores. No significant *F*s for slope were found.

DISCUSSION

The present study represents an attempt to isolate two sources of variance in perceptual recognition scores: variance related to highly overlearned response habits and variance related to the presence of a visual stimulus. The results indicated that *S* does enter the perceptual situation with clearly defined response habits which are not under the control of the perceptual stimulus and which can influence *S*'s recognition score. With regard to the perceptual defense effect words classified as anxiety arousing have less probability of being used as recognition guesses and this negative response bias makes a correct signal detection less likely when a discriminative stimulus is presented.

Comparison of Groups SA and SP did not indicate that the presence of a perceptual stimulus produced a perceptual defense effect greater than that expected on the basis of response bias alone. Thus, the present research supports a response probability theory of perceptual defense.

It should be noted that the present

method of isolating response variance from perceptual variance does not exhaust the methods for attacking this problem. Despite the absence of a significant difference between Groups SA and SP in the magnitude of the perceptual defense effect, it is possible that different sets of variables could be producing the same effect in the two groups. More conclusive evidence for the exclusive role of response bias in producing the perceptual defense effect will come from studies which find a means for equalizing the probability of usage of the two classes of indicator responses.

SUMMARY

The role of response factors in producing the perceptual defense effect was evaluated. Sixty *S*s were assigned to three groups: Group SA (Stimulus Absent), Group SP (Stimulus Present), Group II (Increasing Information). Stimuli, presented tachistoscopically, were pairs of words (half anxiety arousing, half neutral in meaning) selected from *S*'s word association protocol.

Results were as follows: (*a*) A negative response bias (for anxiety-arousing words) is present when no discriminative stimulus is present. (*b*) This negative response bias would be sufficient to produce significant differences in accuracy, as estimated by pseudoaccuracy scores. (*c*) Comparison of pseudoaccuracy scores from Group SA with accuracy scores from Group SP indicates that presence of a discriminative stimulus did not produce a perceptual defense effect greater than would be expected by response bias alone.

The results are interpreted as supporting a response probability theory of perceptual defense. It is suggested that final evaluation of this theory awaits studies in which perceptual recognition scores are obtained under conditions in which no response bias is possible.

REFERENCES

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious processes: An analysis in terms of psychophysical indicator methodology. *Psychol., Bull.*, 1958, **55**, 373–411.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

(Received October 1, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 29–31

PREDICTING INSTRUMENTAL PERFORMANCE FROM THE INDEPENDENT RATES OF CONTINGENT RESPONSES IN A CHOICE SITUATION[1]

AARON J. BROWNSTEIN[2]
University of Missouri

The present experiment derives from a theory of positive reinforcement (Premack, 1959, 1961) in which the necessary and sufficient condition for reinforcement is a rate differential. That is, if a response of higher rate is made contingent upon the occurrence of a response of lower rate, then an increment in the rate of the lower rate response will obtain. The determination of higher and lower rate response is carried out in a noncontingency, sometimes called operant level, situation. (See Premack, 1959, 1961, for a more detailed discussion of this procedure.) The rate so obtained is called independent rate. Thus, the theory attempts to predict rate in the contingency situation, called dependent rate, from the independent rates.

The theory assumes that the increment given the lower rate response is proportional to the independent rate of the contingent response. Furthermore, independent rates of like numerical value are of like reinforcement value, irrespective of the responses in which the rate is manifested or the parameters with which the rate is obtained.

The theory contains no special provisions for predicting behavior in a choice situation. Dependent rates in the choice situation should be proportional to independent rates of the contingent responses, as measured *in the choice situation.* Thus the theory recognizes the choice situation only as a possible parameter of the independent rate of the contingent responses.

Within the present framework prechoice dependent rate differences would be expected to be maintained, modified, or reversed in the choice situation only as the introduction of choice maintains, modifies, or reverses the difference between the independent rates of the contingent responses. This approach differs radically from approaches in which instrumental response strength for the subsequent components of the choice are made the basis for predicting behavior in the choice situation.

Young and Greene (1953) have shown that, when *S*s are permitted to drink sucrose concentrations with one concentration present, volume consumed is inversely related to sucrose concentration. When a choice of concentration is permitted *S*, volume consumed is directly related to concentration. In that study, no change in amount consumed accompanied the transition from single-concentration tests to choice tests for the higher concentration. However, drinking of the lower concentration was drastically reduced in the choice tests.

On the assumption that lick rate is positively related to volume consumed, the predictions which follow from Premack's (1959, 1961) position are quite clear. First, when only one concentration is present, dependent rate will be inversely related to the concentration. Second, when a choice of concentration is available, dependent rate will be positively related to concentration. Finally, in the choice situation, dependent rate for the higher concentration will be equal to the dependent rate for this concentration when tested singly, whereas dependent rate for the lower concentration will be markedly reduced.

---

[1] This investigation was supported by Research Grants M-817 and M-3345 from the National Institute of Mental Health, of the National Institutes of Health, Bethesda, Maryland.

[2] Now at Florida State University.

METHOD

Subjects.—The *S*s were four naive female albino rats approximately 110 days old at the start of the experiment.

Apparatus.—The apparatus used was a Skinner box with inside dimensions of 8¾ × 6 × 7½ in. The four sides and top of the box were constructed of ¼-in. Masonite. The floor of the box was ½-in. hardware cloth. Two levers, which could be presented singly or together, were mounted on one wall of the box 3 in. above the floor. The levers were 2½ in. apart, separated by a ¼-in. piece of Plexiglas, and protruded ¾ in. into the box. Two liquid reinforcement magazines and two dippers delivered sucrose concentrations through openings in the floor of the box beneath the lever pressed. Each lever press activated only the dipper beneath the lever pressed.

Procedure.—One week prior to the start of the experiment proper, all *S*s were placed on a 23-hr. food deprivation schedule that was maintained throughout the experiment.

Two days of magazine training were administered. Each session was 40 min. long and consisted of 40 presentations of each dipper in random order. For each *S* one dipper presented 8% sucrose concentration and the other 32% throughout the experiment. Concentration was counterbalanced over *S*s with respect to position of the lever.

Twenty sessions of single-lever pressing were then given. Each session was 20 min. in duration and each press reinforced. An ABBA order was employed in determining which lever and associated concentration was presented during a session. The intersession interval was 24 hr. for the first nine sessions, 48 hr. for all remaining sessions.

Following single-lever sessions, seven 20-min. sessions with both levers and both associated concentrations were administered.

TABLE 1

ANALYSIS OF VARIANCE OF MEAN BAR PRESSES PER SESSION

Source	*df*	*MS*	*F*
Concentration (C)	1	3329.29	5.61*
Type of session (T)	1	18975.06	31.98**
C × T	1	22455.71	37.84**
Error	12	593.37	

* $P < .05$.
** $P < .001$.

RESULTS AND DISCUSSION

The mean number of lever presses for each *S* on the last five single-lever sessions of each concentration and the mean number of presses on each lever for the seven double-lever sessions were determined. An analysis of variance was then carried out on these data. The summary of the analysis presented in Table 1 shows that the interaction between concentration and type of session is highly significant ($P < .001$). Figure 1 presents a plot of this interaction. As was predicted, the dependent rate for the lower concentration was greater than that for the higher concentration in single-lever tests. In double-lever tests, however, the rate for the lower concentration was drastically reduced

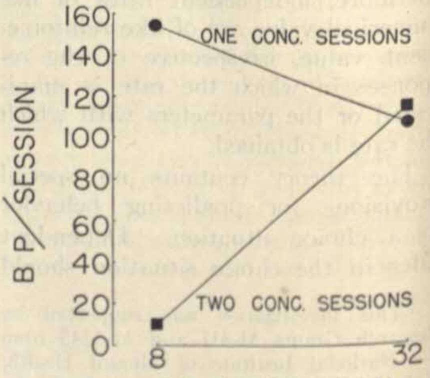


FIG. 1. Bar pressing as a function of concentration and type of session.

(approximately 1/15 of that obtained in single-lever tests). The rate on the higher concentration is approximately equal in both types of tests.

In support of the analysis, examination of individual data shows that for 3 of the 4 *S*s during the five single-lever sessions the lowest rate obtained on 8% contingent pressing was higher than the highest rate obtained on 32% contingent pressing. On each of the seven double-lever sessions the lowest rate obtained on the 32% contingent pressing was higher than the highest rate obtained on the 8% pressing for all *S*s.

The results of this study clearly support Premack's (1959, 1961) assumptions regarding the independent rate of the contingent response as a determiner of the dependent rate. An unequivocal relationship between independent rate of contingent response and dependent rate was demonstrated in that reversal of the independent rate of the contingent response was paralleled by the reversal in the dependent rate.

Recently, Premack (1961) has demonstrated a reversal in dependent rate accompanying a reversal in the independent rate of drinking on various sucrose concentrations where the parameter of concern was duration of licking prior to introduction into the contingency situation. As has been noted (Premack, 1961), demonstrations with the consummatory response contribute little to testing the assumption that a rate differential is a necessary condition for reinforcement. Nevertheless, support for the sufficiency of the condition accrues as the parameters governing the rate of the consummatory response function are shown to yield congruous consummatory and dependent rate functions.

Finally, returning to the parameter of concern in the present study, the traditional approach to this problem has been to ascertain the bar-press rates in the single-bar situation and assume some mode of interaction, e.g., ratio, subtractive, etc., to predict rate in the choice, or conflict, or competition, or selective learning situation. It is clear that no interaction rule as yet offered could predict the superiority of the 32% contingent bar. Since no change in the variables determining the prechoice response strengths, maintenance schedule, reinforcement magnitude, etc., accompanied the transition to the choice situation, the direction of the prechoice difference would be expected to be maintained.

SUMMARY

Four *S*s were given bar-press tests with 8% and 32% sucrose concentrations presented singly and simultaneously. When tested singly, bar-press rate was inversely related to concentration. When tested simultaneously, a direct relationship was obtained. The direct relationship is due to the much suppressed responding on the 8% contingent bar with no change in rate of the 32% contingent bar. The fact that the bar-press rates parallel rates obtained for drinking in single and simultaneous tests supports Premack's (1959, 1961) theory.

REFERENCES

PREMACK, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychol. Rev.*, 1959, 66, 219–233.

PREMACK, D. Predicting instrumental performance from the independent rate of the contingent response. *J. exp. Psychol.*, 1961, 61, 163–171.

YOUNG, P. T., & GREENE, J. T. Quantity of food ingested as a measure of relative acceptability. *J. comp. physiol. Psychol.*, 1953, 46, 288–294.

(Received October 5, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 32–35

MAGNITUDE OF REWARD AND PROBABILITY LEARNING[1]

YVONNE BRACKBILL, MICHAEL S. KAPPY, AND RAYMOND H. STARR

Johns Hopkins University

A first concern in studies of probability learning—and still a persistent one—is the "matching" choice behavior that *S*s show when confronted with events of unequal occurrence. Instead of behaving like rational game theorists by learning to respond at unity, *S*s tend to match the relative frequencies of their responses to the relative frequencies of the stimulus events (Grant, Hake, & Hornseth, 1951). Goodnow (1955) has advanced an interesting set of hypotheses concerning factors that lead to matching; she cites intolerance for boredom, inability to overlook losses, and so on. Statistical learning theories provide a more formal approach to the prediction of probability matching (cf. Bush & Estes, 1959).

In terms of experimental evidence, one variable that appears to be of no small importance in predicting and manipulating relative response frequencies is the nature of tangible reinforcing contingencies offered by the experimental procedure. First, the mere introduction of reward for correct predictions has been shown to increase maximum gain responding. Money may not be everything, but it can do a lot toward alleviating the boredom that comes from making the same verbal response for 100 trials or more (Siegel & Goldstein, 1959). Further, it has been found that maximum gain responding is increased by the use of both reward for correct prediction and punishment for incorrect prediction (as in the "risk" condition of Siegel & Goldstein, 1959), by a previous history of infrequent rewards (or low "expectancy for success," Stevenson & Zigler, 1958), and by higher frequencies of reward in a contingent procedure (Stevenson & Weir, 1959).

In the present study, the effect of amount of reward was investigated, with the expectation that *S*s will maximize gain to the extent that there is something tangible to be gained by so doing. In other words, the hypothesis was that in the typical noncontingent probability learning situation, maximum gain strategy is a direct function of the amount of reward offered for correct predictions.

METHOD

Subjects.—The *S*s were 24 boys and 24 girls from the second grade classes of a single elementary school. Mean age was 95.5 mo., $SD = 9.4$. Mean IQ was 85.8, $SD = 18.9$, by the SRA test. The sample was relatively homogeneous in terms of socioeconomic background; modal occupation level for fathers corresponded to Class 5 of Warner's (1949) 7-point scale. Thirty-four *S*s were white, 13 were Negro, and 1 was Oriental.

Experimental material.—This consisted of 3 × 5 in. cards cut from green, light weight cardboard stock. On each card was printed the picture of either a cat or a dog. Two packs of 200 cards each were prepared, in order that dog or cat would appear equally as the more frequent event over treatment levels, sex of *S*s, and sequence of running *S*s. In preparing these packs, the relative proportions of the two stimulus events were held constant for all treatment levels at .75:.25. A randomized series of events was prepared (and recorded) in blocks of 20 trials each, with the overall objective probability split of .75:.25 maintained within each block. There was no other restriction on randomization. Twenty such 20-trial blocks were prepared. Then, for each *S*, a 10-block sequence was randomly drawn from the 20-block pool.

---

[1] This investigation was carried out during the term of a Postdoctoral Fellowship awarded to the first author by the National Institute of Mental Health. The authors are grateful for the cooperation of Mary Adams, Dorothy Rawlings, and the staff of the Barclay Elementary School, Baltimore, Maryland.

In addition to the pack of 200 cards, a box containing approximately 750 marbles and a cup with a 100 marble capacity in which *S* accumulated his winnings were also on the table at which *E* and *S* worked. On display in another part of the room was a standard array of ten 10¢ toys. These were high-preference toys, selected on the basis of previously obtained rankings from same-age *S*s.

Procedure.—Forty-eight *S*s were randomly assigned to four experimental groups, with the single restriction that each group of 12 contain equal numbers of boys and girls. The four groups were given 5, 3, 1, or 0 units of reward (marbles) for each correct response. One hundred marbles could be exchanged for one toy.

At the beginning of a series, the pack of 200 cards was face down; *S* was instructed to guess before *E* turned each card over, whether the card would show a dog or a cat. Both at the beginning of the experiment and subsequently, whenever he had accumulated 100 marbles, *S* was asked if he wanted to try to win a toy (or another toy). If he responded negatively at any choice point, the procedure was discontinued. One prospective *S* was discarded at the beginning of the experiment after stating that none of the available toys interested her.

RESULTS AND DISCUSSION

Figure 1 shows the mean proportions of predictions of the more frequent event ($S_{.75}$) in successive 20-trial blocks for the four groups. As a test of the hypothesis that maximum gain responding increases as a function of magnitude of reward for correct predictions, the Jonckheere (1954) test against ordered alternatives was applied to the data from the last three blocks. The resulting statistic of 2.65 is significant at the .004 level by a one-tailed test. (Scores from three blocks were used in order to minimize the number of tied scores. If one used the same test on scores of the last block only and split the ties, the resulting statistic of 1.83 would still be significant at the .02 level.)

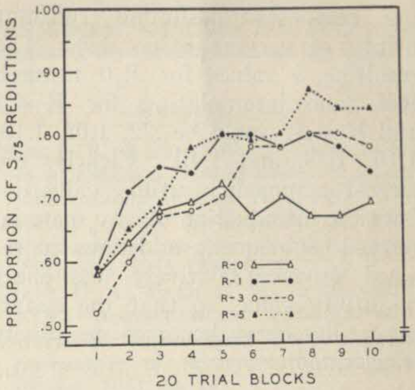


FIG. 1. Mean proportions of $S_{.75}$ predictions in successive 20-trial blocks for four experimental groups differing in number of units of reward offered for every correct prediction. (Each plotted point is a Group × Block mean based on 12 cases.)

Inspection of Fig. 1 shows that, for the last half of the trials, the most noticeable difference among treatment levels is between nonreward and any value of reward. This suggests that the performance function under amount of reward is not linear for these data. The same suggestion has been made by Pubols (1960), who concluded from his review of studies on amount of reward in learning and performance in animals, that the performance function ". . . is negatively accelerated, and possibly logarithmic . . ." (p. 111). If one prefers a technique more sophisticated than simple visual inspection, the same question may be considered from the standpoint of utility theory. On the basis of Siegel's (1959) Model I (Equation 1), it is possible to calculate the ratio of the utility of making correct responses to that of varying responses (assuming constant utility of varying responses). The resulting α values for R-0 through R-5 (and interpolating for R = 2 and R = 4) would be .72, 1.08, 1.13, 1.16, 1.29, and 1.40. Clearly, the largest increment in utility value follows the introduction of any material reward; subsequent additions in reward produce relatively less effect on utility value, so that the resulting utility curve becomes negatively accelerated.

Sequence Analysis

A first-order sequence analysis was performed on the last 100 trials for each *S*. Such an analysis yields the probability of *S*'s prediction on Trial *n*, given two items of information about Trial *n*-1: (*a*) the event that *S* had predicted previously, and (*b*) the event that had actually occurred previously. Since the Jonckheere (1954) analysis had already shown tangible rewards to have a significant effect on responding, it is reasonable to expect that previous actual occurrence rather than previous prediction would emerge from a sequence analysis as the more dependable predictor from Trial *n*-1 to *n*. This expectation finds support in a comparison of Columns A and B, Table 1; the average difference between the two sets of probabilities is a mere .04. Further, the ordering within the two sets of probabilities reflects with fair accuracy the effect of the experimental variable, magnitude of reward.

However, comparison of Columns C and D reveals a different outcome. Prediction to Trial *n* is not a simple function of previous actual occurrence on Trial *n*-1, as it appears to be for the A-B comparison. In fact, the mean difference between the two sets of 48 correlated proportions is significantly different from zero ($CR = 2.82$, $P < .01$). This means that, for those trials on which the less frequent event actually occurred, not only that occurrence but also *S*'s own previous prediction determined his next prediction. Specifically, *S* was much more willing to revert to a maximum gain strategy on his next prediction *if* he had just successfully predicted the less frequent event than if he had not predicted it and it had occurred. This effect appears to be independent of the experimental variable, since there is no systematic trend among C-D difference scores as a function of treatment level.

What is the interpretation of this effect, and how does it relate to the experimental variable? Obviously, the best

TABLE 1

PROBABILITY OF PREDICTING $S_{.75}$ ON TRIAL *n*, GIVEN CERTAIN INFORMATION ABOUT TRIAL *n* − 1

Group	Event on Trial *n* − 1			
	A *S* Predicted $S_{.75}$ & $S_{.75}$ Occurred	B *S* Predicted $S_{.25}$ & $S_{.75}$ Occurred	C *S* Predicted $S_{.25}$ & $S_{.25}$ Occurred	D *S* Predicted $S_{.75}$ & $S_{.25}$ Occurred
R-0	.74	.82	.56	.38
R-1	.80	.83	.85	.65
R-3	.82	.87	.73	.67
R-5	.89	.87	.78	.64

Note.—Each entry represents the mean of 12 individual proportions.

strategy to maximize tangible reward is for *S* to predict the more frequent event on every trial. Further, this strategy is the more compelling the more the reward. That *S*s do behave in accordance with this strategy was shown by the Jonckheere (1954) analysis. What the sequence analysis shows is that there is a second source of reinforcement at work: the utility to *S* of correctly predicting the occurrence of the less frequent event. Where the investigators had intended only one game, *S*s discovered two.

Whether or not one agrees with this interpretation, the fact remains that the net result of this secondary effect is to subtract from maximum gain responding. In effect, it constitutes a response that is competing with the criterion response. However, in contrast with the results of studies in which incentive magnitude has been found to affect both criterion response and competing responses (Pereboom & Crawford, 1958), the present results show only the criterion response to have been influenced by magnitude of reward. The competing response appears to have been maintained by some other form of reinforcement, independent of the form under study.

SUMMARY

This study investigated probability learning in a binary choice, noncontingent procedure with 0, 1, 3, and 5 units of reward for each correct prediction, 100 units being exchangeable for one toy. The *S*s were second-grade children. Maximum gain responding increased with larger values of reward, although the function relating these two variables did not appear to be linear. From a first order sequence analysis, an independent effect emerged through an unexpected source of reinforcement: the utility to *S* of correctly predicting the occurrence of the less frequent event.

REFERENCES

BUSH, R. R., & ESTES, W. K. (Eds.) *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959.

GOODNOW, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

JONCKHEERE, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

PEREBOOM, A. C., & CRAWFORD, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56**, 82–85.

PUBOLS, B. H., JR. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

SIEGEL, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

STEVENSON, H. W., & WEIR, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.*, 1959, **57**, 403–412.

STEVENSON, H. W., & ZIGLER, E. F. Probability learning in children. *J. exp. Psychol.*, 1958, **56**, 185–192.

WARNER, W. L. *Social class in America.* Chicago: Science Research Associates, 1949.

(Received October 10, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 36–41

THE INFLUENCE OF STRENGTH OF DRIVE ON FUNCTIONAL FIXEDNESS AND PERCEPTUAL RECOGNITION [1]

SAM GLUCKSBERG [2]

New York University

The present study represents an attempt to conceptualize the functional fixedness phenomenon (e.g., Adamson, 1952; Duncker, 1945) in terms of an S-R model. Historically, this phenomenon has been represented as reflecting perceptual processes. For example, when *S* uses a matchbox as a platform in order to solve the problem of mounting a candle on a vertical screen (Adamson, 1952), he is described as being able to perceive the box as a platform. Failure of problem solution is interpreted as indicating that *S* can only perceive the container properties of the box and not its platform properties.

In a certain sense (Lachman, 1960) the model used to represent behavior is arbitrary. The same functional fixedness problem can be conceptualized in terms of competition between S-R habits, using the box as a container and using it as a platform.

A number of studies conducted within the S-R orientation (Maltzman, Fox, & Morrisett, 1953; Spence, Farber, & McFann, 1956; Spence, Taylor, & Ketchel, 1956) have shown that a high drive can impair performance when the correct response is low in the habit hierarchy. One reason for this is that high drive prolongs extinction of the dominant habit (Perin, 1942) and thus retards the correct habit from gaining ascendancy (Kendler & Lachman, 1958). According to this analysis, if it were possible to control the position of the correct habit (using the box as a platform) in the response hierarchy, then it would be possible to control the course of problem solution. When the correct habit is low in the habit hierarchy, a high drive should retard problem solving. In contrast, when the correct habit is dominant, a high drive should facilitate problem solving. Experiment I tests this analysis. Experiment II extends the analysis to perceptual recognition.

EXPERIMENT I

Method

Subjects.—The *S*s were 139 male undergraduates randomly selected from the introductory psychology course at New York University. Eleven *S*s were discarded; 5 reported familiarity with the experimental task, and 6 had weak color vision. These latter 6 *S*s could not discriminate between red and black at an exposure duration of 20 msec., and could not be used in the perceptual recognition study which is reported as Exp. II. The total sample thus consisted of 128 *S*s.

These 128 *S*s served in both experiments. Half the *S*s were initially given the perceptual recognition tests, while the remaining *S*s had the problem solving task administered first. Three to 4 weeks intervened between successive testing sessions.

Experimental design.—A simple 2 × 2 factorial design was employed involving two levels of drive and two different response hierarchies in a Duncker-type (1945) candle problem. In this problem *S*s are instructed to mount a candle on a wall. Available to them are cardboard match and thumbtack boxes, matches, and thumbtacks. The problem can be solved by tacking an emptied box to the wall and placing the candle in or on it.

---

[1] This paper is based on a PhD dissertation submitted to the Graduate School of Arts and Sciences, New York University. The author is greatly indebted to H. H. Kendler for his guidance and encouragement during the course of the study, and to the Office of Naval Research for support Contract Nonr 0285(30), H. H. Kendler, principal investigator.

[2] Now at United States Army Ordnance Human Engineering Laboratories, Aberdeen Proving Ground, Maryland.

Drive was manipulated by offering different incentives and the response hierarchies were controlled by the manner in which the problem was arranged. When the boxes are presented empty, the correct response is initially high in the hierarchy (hereafter referred to as Cond. Dm+), and the problem is easily solved (Adamson, 1952; Duncker, 1945). When the boxes shown to *S*s are filled with tacks and matches, the correct response is initially low in the hierarchy (Cond. Dm−), and relatively few *S*s succeed in solving the problem.

Experimental materials.—The following materials were presented on a table in standard positions in front of *S*: a round thumbtack box, approximately 50 thumbtacks, a rectangular pocket-size matchbox, approximately 35 safety matches, and one wax candle. A large box was used to conceal the materials while *S* received instructions. A corrugated paper board, 36 × 24 in., was mounted on the plaster wall of the laboratory. This board was removable, and served as the "wall" in the experiment. Each *S* was provided with fresh, unmarked materials.

Procedure.—The *S*s were seated at the table and given instructions on the requirements of the problem. In addition, low-drive *S*s were told: "We are doing pilot work on various problems in order to decide which will be the best ones to use in an experiment we plan to do later. We would like to obtain norms on the time needed to solve."

In place of the above instructions given to low-drive *S*s, high-drive *S*s were told: "Depending on how quickly you solve the problem you can win $5.00 or $20.00. The top 25% of the *S*s in your group will win $5.00 each; the best will receive $20.00. Time to solve will be the criterion used."

When *S* had indicated that he understood the instructions the covering box was removed, *E* said, "begin" and a stop watch was started. Timing was stopped when *S* had either pushed a tack through an empty box or had failed to solve within 15 min. A score of 15 min. was assigned to failure *S*s.

Results and Discussion

The response measure used to evaluate the effects of drive level and response hierarchies on problem solving was time to solve, in minutes. The mean solution times, together with number of failures per group, are presented in Table 1. The results provide evidence consistent with the experimental predictions. When the correct response is dominant (Dm+) high drive tends to facilitate problem solving; when an incorrect response is dominant (Dm−) high drive tends to retard problem solving.

Even though solution time scores were skewed, analysis of variance and *t* tests were applied to the solution time scores (Boneau, 1960). The significant difference ($F = 28.42$, $df = 1/124$, $P < .001$) between Groups Dm+ and Dm− validates the experimental manipulation controlling the position of the correct and incorrect responses in the hierarchy. The Dm+ form of the problem, as expected, is easier than the Dm− form. The difference between the means of the over-all high- and low-drive groups was not significant. On the other hand, the predicted interaction between drive levels and nature of the dominant response was obtained ($F = 7.37$, $df = 1/124$, $P < .01$).

Differences between the means of high- and low-drive groups within response hierarchy conditions were

TABLE 1

SOLUTION TIME SCORES TO SOLVE CANDLE PROBLEM: EXP. I

Group		Number of Failures	Solution Time (Min.)	
Nature of Dominant Response	Drive Level		Mean	*SD*
Dm+	High	2	3.67	4.52
	Low	6	4.99	5.42
Dm−	High	16	11.08	4.84
	Low	10	7.41	5.63

evaluated by *t* tests. The difference was significant ($t = 2.73$, $P < .01$) in Cond. Dm− but not in Cond. Dm+.

The failure to obtain a significant effect in Cond. Dm+ may be attributed, in part, to the fact that the performance of the correct response may have been too easy, producing a ceiling effect. Drive, as an energizer, is expected to increase speed of responding. The correct response, when it occurred, was executed fairly rapidly, usually within 5–10 sec. The effect of drive upon solution time was thus limited by the simplicity of the behavior required. Also, the response hierarchies for all *S*s may not have been adequately controlled. The facilitative effect of drive could well have been masked because the correct response was not dominant for some *S*s. For some *S*s this was obviously the case; 8 of the 64 *S*s in Group Dm+ failed to solve (see Table 1). Some *S*s who did solve the problem attempted initially to melt wax and use the wax as an adhesive. This response, of course, is incorrect. If the dominant response is not the correct response, as the above suggests, then drive should not facilitate problem solving. Thus the lack of significance might be due to the fact that the results obtained from *S*s whose dominant response was truly the correct response were offset by the results of *S*s whose initial responses were incorrect. It would seem, if one is to judge by the above, to be extremely difficult to arrange a problem situation in which all *S*s respond initially by making a correct response.

EXPERIMENT II

A representation of functional fixedness within S-R drive theory (Spence, 1956) proved to be fairly accurate. The problem that is now posed is whether the same representation can be applied to a simple perceptual task. If the S-R model applies to problem solving behavior that has been traditionally represented in perceptual terms, then can the same model also represent a perceptual task itself? A perceptual task (word recognition) was designed to be analogous to the problem solving task. The effects of drive were investigated under two conditions: dominant response correct (Dm+), and incorrect (Dm+).

TABLE 2

STIMULUS WORDS USED IN WORD RECOGNITION TASK: EXP. II

Dm+		Dm−	
PAINTER	CHURCH	ELEGIES	VIGNETTE
SAVINGS	SERVICE	FRUGALITY	AMICABLE
LAWYER	HEAVEN	ASSIDUOUS	CHANCELS
POETRY	MARKET	PERCIPIENCE	MUNDANE
EDUCATION	DEBATING	ERUDITION	UNCOERCED
PHYSICS	SYMPATHY	BEATIFIC	BENIGN

Method

Subjects.—The 128 *S*s of Exp. I served in this study.

Experimental design.—The same 2 × 2 factorial design in Exp. I was used. A word recognition test was administered under one of two drive conditions (high or low). The task required either the recognition of an obvious word or one that was difficult to identify. The *S*s confronted with the former task were members of the Dm+ groups. The latter belonged to the Dm− groups.

The words to be recognized were presented tachistoscopically. Two words were presented simultaneously on each trial. One word, high on the Thorndike-Lorge (1944) word count, was printed in black upper case letters, and was easy to recognize. The other was low on the Thorndike-Lorge word count, printed in pale red lower case letters, and was difficult to identify (Howes & Solomon, 1951).

Twelve such word pairs were presented. The words used are contained in Table 2. The *S*s in the Dm+ groups were instructed to identify the black words while *S*s in the Dm− groups were required to identify the red words. The *S*s were instructed as follows:

> This is a tachistoscope. If you look into the eyepiece you will see a black cross on a white ground. That is your fixation point; any stimuli that appear will appear at that distance from your eyes in just about that general area. I am going to flash words on the screen. Each stimulus card will

> have two words printed on it. One of the words will be printed in black, one in red. I am gradually going to increase the exposure time until you can tell me what the red (or black) word is. Remember, we are only interested in the red (black) word. Try to report the word as soon as you can. Wrong guesses will not be counted.

The *S*s were assigned to one of the four experimental groups in the following manner. The 64 *S*s whose initial experimental task was the perceptual recognition problem were randomly distributed equally among the four groups. Subsequently, the 16 *S*s in each of these four groups were proportionally assigned, at random, to the four problem solving groups. The 64 *S*s who were initially given the candle problem were assigned to the experimental groups of this study in a similar fashion. This procedure minimized systematic sequence effects and permitted separate statistical treatment of the two experiments.

Apparatus.—The 12 word pairs were presented in a modified Dodge tachistoscope, described in detail by Karlin (1955). This instrument permits exposure durations ranging from .001 sec. to 2 min.

Procedure.—The *S*s received instructions as described above, and were then differentiated into high- and low-drive groups by instructions similar to those given in Exp. I, i.e., high-drive *S*s were offered money, low-drive *S*s were not.

When the instructions had been given, visual duration thresholds were taken for each of the 12 words by the ascending method of limits, in 5-msec. steps. A current of 20 ma. was supplied to the field and target lamps.

The procedure followed was that described in detail by Howes and Solomon (1951) with the following exceptions: (*a*) the criterion for correct recognition was one correct verbal report instead of three consecutive reports; (*b*) since the thresholds for the individual words were of no interest the word order for all *S*s was the same; and (*c*) no practice trials were given. The score for each *S* was the mean visual duration threshold for all 12 words.

Results and Discussion

The mean and median thresholds (in msec.) are presented in Table 3. All differences obtained between groups were in the predicted directions. In Cond. Dm+ high-drive *S*s were superior to low-drive *S*s. The direction of the difference was reversed in Cond. Dm−. Because of the marked heterogeneity in the group variances, an analysis of variance was made on log transformed scores. The means and *SD*s of the transformed scores, together with the relevant *t* tests, are presented in Table 4.

TABLE 3

MEAN AND MEDIAN VISUAL DURATION THRESHOLDS (MSEC.)

Group		Visual Duration Thresholds			
Nature of Dominant Response	Drive Level	Mean	*SD*	Median	Range
Dm+	High	34.0	14.2	29.5	17.0– 75.0
	Low	47.0	27.1	41.0	14.0–145.0
Dm−	High	199.8	69.8	207.5	77.0–347.0
	Low	151.9	83.7	133.0	42.0–430.0

The significant difference between Groups Dm+ and Dm− ($F = 330.70$, $df = 1/124$, $P < .001$) serves to validate the experimental manipulation of response hierarchies, i.e., Cond. Dm+ is easier than Cond. Dm−.

As expected, the difference between the over-all means of the high- and low-drive groups was not significant.

TABLE 4

MEANS, VARIANCES, AND *t* TESTS OF LOG$_{10}$ VISUAL DURATION THRESHOLDS

Group		Log Visual Duration Thresholds		
Nature of Dominant Response	Drive Level	Mean	S^2	t[a]
Dm+	High	0.4909	0.8658	2.46*
	Low	0.6154	1.6407	
Dm−	High	1.2708	0.9003	2.89**
	Low	1.1234	1.6054	

[a] Mean square of within group variance used as estimate of population variance.
* $P < .02$.
** $P < .01$.

The predicted interaction between drive level and nature of the dominant response was significant ($F = 14.30$, $df = 1/124$, $P < .001$). The differences between the means of transformed scores of high- and low-drive groups within response hierarchy conditions were evaluated by *t* tests (see Table 4). The high-drive *S*s have significantly lower thresholds under Cond. Dm+ ($P < .02$) and significantly higher thresholds under Cond. Dm− ($P < .01$).

The question arises, what would have happened if, in Cond. Dm−, the red word had been presented alone? The recognition of the black word, including attending to it, was assumed to be the incorrect dominant response. However, perceiving the black word is not the only incorrect response possible for Group Dm−. The initial response to the red word can also be incorrect, and when it is it must undergo extinction. The *S*s in the Dm− groups consistently misreported the red words: BEATIFIC was given as BEAUTIFIC or BEAUTIFY, ELEGIES as ELEGANT, CHANCELS as CHANNELS, etc. Furthermore, the errors tended to persist over several trials, in one case for 20 trials until the word (CHANCELS) was correctly identified.

The presence of the black word, then, was probably unnecessary. That is, the same results would have been obtained for Group Dm− if only the red words had been exposed. Since the dominant response (the misreport of the word) would still be incorrect, a high drive should prolong its extinction and thus impair correct recognition. Following this line of argument, it would be predicted that there is an interaction between drive level and word frequency. If an inverse relationship is postulated between word frequency and probability of misreport (i.e., as word frequency decreases the probability of erroneous report increases), then a high drive will impair recognition of low frequency words. Conversely, a high drive will facilitate recognition of high frequency words when the probability of error is low.

In conclusion it should be noted, first, that neobehavioristic drive theory has been successfully extended to two different areas. In each case the major prediction of the theory was confirmed: there is an interaction between drive level and nature of the response hierarchy. Secondly, the extension of the theory could be made only by representing functional fixedness and perceptual recognition in S-R terms. Once this had been done a limited number of principles derived from conditioning and simple selective learning sufficed to account for the obtained effects of drive. Of course, the phenomena concerned are not to be equated with conditioning, even though the conditioning model proved useful (Lachman, 1960). Instead, it is suggested that important empirical consequences have been obtained, indicating the utility of an S-R approach to problem solving.

SUMMARY

A problem solving task and a perceptual recognition task were represented in S-R terms. Predictions derived from neobehavioristic drive theory were then tested in each situation. Each task was presented under one of two drive levels (high or low) and in one of two forms (dominant response correct or incorrect). High drive impaired performance in both tasks when the dominant response was incorrect. When the dominant response was correct high drive facilitated performance in the perceptual recognition task. The results in the problem solving task, although in the predicted direction, failed to achieve statistical significance. The major prediction of the theory, that there is an interaction between drive level and nature of the dominant response, was confirmed.

REFERENCES

ADAMSON, R. E. Functional fixedness as related to problem solving: A repetition of three experiments. *J. exp. Psychol.*, 1952, **44**, 288–291.

BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, **57**, 49–64.

DUNCKER, K. On problem-solving. *Psychol. Monogr.*, 1945, **58**, (5, Whole No. 270).

HOWES, D. H., & SOLOMON, R. L. Visual duration thresholds as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.

KARLIN, L. The New York University tachistoscope. *Amer. J. Psychol.*, 1955, **68**, 462–466.

KENDLER, H. H., & LACHMAN, R. Habit reversal as a function of schedule of reinforcement and drive strength. *J. exp. Psychol.*, 1958, **55**, 584–591.

LACHMAN, R. The model in theory construction. *Psychol. Rev.*, 1960, **67**, 113–129.

MALTZMAN, I., FOX, J., & MORRISETT, L., JR. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46**, 50–59.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received October 11, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 42–49

PREFERENCE PROBABILITY BETWEEN GAMBLES AS A STEP FUNCTION OF EVENT PROBABILITY[1]

R. DUNCAN LUCE AND ELIZABETH F. SHIPLEY
University of Pennsylvania

This experiment is another of the studies arising ultimately from the utility ideas of Ramsey (1931) and von Neumann and Morgenstern (1947). It is based on a probabilistic analysis of choices among uncertain outcomes that rests upon two major assumptions. (Luce, 1959, Ch. 3). One, a general choice axiom, relates the choice probabilities for any set of three alternatives to those for its two-element subsets. Specifically, it postulates that if one of the three alternatives is eliminated, its probability is distributed between the other two in proportion to their probabilities in the original set of three. The second, a decomposition axiom (Luce, 1958; Luce & Raiffa, 1957) which has meaning only for uncertain outcomes, assumes that when *S* is choosing between certain pairs of gambles his preferences between the pure outcomes are statistically independent of his likelihood judgments between the chance events. More exactly, it says that if one knows the objective probability that one pure outcome is preferred to another and the objective probability that one event is deemed more likely to occur than another, then the probability of choosing between the gambles generated by these outcomes and events is given by the straightforward probability calculations under the assumption of independence. Although nothing is assumed in this postulate about the existence of subjective probability or utility functions as in classical utility theory, it is of the same spirit as the expected utility hypothesis, but much weaker. The precise mathematical statements of these two assumptions can be found in Luce (1959); they will not be repeated here.

Some psychophysical evidence supports the choice axiom (Anderson, 1959; Clarke, 1957); however, as yet, no direct preference tests have been made. Coombs (1958, 1959) presents indirect negative evidence, but there is some doubt about its meaning (Luce, 1959, p. 19). No data have yet been published on the decomposition axiom. On both counts, therefore, data are needed. Because these assumptions are difficult to test separately, and because together they imply a surprising result, we shall test only their conjunction.

Consider the one-person game:

$$\begin{array}{ccc} & \text{Option I} & \text{Option II} \\ \rho & \left[a \right. & \left. c \right] \\ \bar{\rho} & \left[b \right. & \left. d \right] \end{array}$$

where a, b, c, and d are sums (possibly negative) of money such that $a > c > d > b$ and where ρ and $\bar{\rho}$ mean the occurrence and nonoccurrence, respectively, of a certain chance event. The *S* selects one of the two options and then, following his choice, the chance event is given an opportunity to occur. These two decisions—one by *S*, one by chance—determine a cell in the matrix, which contains *S*'s payoff. Thus, if *S* chooses Option II and the chance event fails to occur (row $\bar{\rho}$), the payoff to *S* is d.

---

[1] This research was supported in part by Grants NSF G-4506 and G-5544 from the National Science Foundation to Harvard University. The experimental work was carried out at the Laboratory of Social Relations, Harvard University.

Now, holding the payoffs fixed, consider *S*'s choices as we vary the event probability. If the event is certain to occur—or so *S* believes—then it is only sensible for him always to choose Option I because chance will choose Row 1 and, by assumption, *S* prefers *a* to *c*. On the other hand, if the event is certain not to occur, *S* surely will never choose Option I because $b < d$. And as the event probability is varied from zero to one, *S*'s probability of choosing Option I must change from zero to one. The question is: what is the nature of this transition?

Classical economic thought has assumed a simple discontinuity, but most psychologists doubt that this is correct, and the evidence (Chipman, 1960; Coombs, 1958, 1959; Davidson & Marschak, 1959; Davidson, Suppes, & Siegel, 1957; Mosteller & Nogee, 1951; Papandreou, 1957), although open to other interpretations, is generally viewed as being negative. Assuming that the function is not a simple discontinuity, the most plausible a priori conjecture is a smooth ogival function, so familiar from psychophysics. Plausible, yes, but not possible for all choices if the two assumptions mentioned above are correct. For payoff matrices where one row dominates the other (e.g., $a > b$ and $c > d$), they imply that the choice probability varies from zero to one by a series of jumps. That is to say, the function must be a step function, of which the simple discontinuity is a special case.[2] When neither row dominates the other, i.e., when the two largest payoffs lie on a diagonal, no theoretical prediction has been derived. It is not known whether steps are again necessary or whether continuous functions are possible.

Although this prediction of steps is qualitatively startling and sharply at variance with most psychologists' intuitions, it is uncertain whether it can be successfully tested—not only would negative results be completely ambiguous, but seemingly positive ones are open to other interpretations. Basically, there are three difficulties. First, the theory states neither how many steps to expect nor where to find them. If there are enough, say 8 or 10, reasonably spaced over a sufficiently narrow region of event probabilities, then a vast amount of data is needed to tell a step function from a smooth ogive. Second, to estimate the choice probabilities, the choices must be presented repeatedly, allowing all the well-known dangers of systematic changes over time. And third, even if by good luck and design these troubles are avoided, one must still live with the inevitable and considerable binomial variance of our observations which, on the one hand, may tend to mask the steps of a step function (unless they are broad) and, on the other hand, may introduce at least small steps into data arising from an ogive.

Severe though these difficulties seem to be, it seemed worthwhile to conduct the obvious experiment. The payoff, should the data hint at positive results, seemed sufficiently large to offset both the slim chance that the steps are there and the inherent uncertainties of knowing it even if they are.

METHOD

Subjects.—Without specifying the exact nature of the work, five Harvard College students were solicited to participate in an experiment for which they would be paid both an hourly wage and a bonus. After he had indicated interest in the work, each was asked if he had qualms about gambling; none did.

Apparatus.—The 2×2 payoff matrices were printed on small white cards. Each column defined a gamble, and so an option

[2] It should be noted that this conclusion, which is "global" in nature, follows from repeated application of Luce's (1959, p. 86) Theorem 13, which is "local" in nature. The theorem requires that various probabilities of choice be different from 0 and 1, which is what gives it its local character, but this assumption need not hold for any two arbitrarily chosen gambles of an experiment designed to test the step-function hypothesis. Thus, there can be no question of looking to the data to see whether prerequisites of the step-function hypothesis are satisfied.

for *S*; each row corresponded to one of two complementary events. Packs of these cards, face down, were mounted in a rack before *S*. Turning a card over to expose the matrix tripped a microswitch which in turn activated a pen of an Esterline Angus recorder. Two buttons, labeled to correspond to *S*'s two options, were located in front of each *S* and were wired to the pen recorder.

The events were specified in terms of the faces shown by five dice. The possible outcomes were ranked into hands, much as in poker except that flushes could not occur, there being no suits, and straights were defined to be of the lowest rank. A hand known as the house's hand was printed on each card. One event was defined to have occurred when the dice exhibited a hand, known as *S*'s hand, higher than the house's; the complementary event occurred when *S*'s hand was the same as or less than the house's. Each *S* was provided with an ordered list of the 252 hands that might arise and of the probabilities of beating each of them. The events were produced by spinning a wire cage that contained the five dice.

Procedure.—Prior to the experiment, *E* explained both the procedure and the bonus system. The payoffs in the matrices were treated as points to be won or lost on each throw of the dice. At the end of the experiment the accumulated points would then be used to determine each *S*'s share of a $200 bonus. They were told that a point was roughly equal to 1 cent, which was true.

An experimental trial was as follows. When *E* gave a ready signal to begin the trial, each *S* turned over the next card in his pile, examined the alternatives, and indicated his choice by pressing one of his two buttons. The time when the card was turned over, his choice, and the time it was made were all recorded. After all *S*s had made their choices, one of them—the same one throughout an experimental session—spun the cage and called out the resulting hand. Each *S* then recorded on a sheet provided the points won or lost on the trial. They knew that on any given trial the payoff matrix and the house's hand differed among them, but that throughout the experiment they would all be presented with the same six matrices. They were also led to believe, although this was not true, that they would all see the same house hands.

Four preliminary sessions were run to determine for each *S*-matrix combination the range of events in which *S*'s choices were uncertain. The experiment itself consisted of daily sessions of 150 trials, lasting about 1 hr., for 33 days. Because some of the initial choices of event ranges for S_1 were inappropriate, he was run three extra sessions. For the first few weeks, sessions occurred only when all 5 *S*s could attend; however, S_1 became ill, which made it necessary to relax this requirement. It seemed that the larger the group, the more enjoyable the session; some sessions at the end with a single *S* were considered dull. On the whole, however, if noisy rivalry is any sign, the *S*s were interested and involved.

TABLE 1

PAYOFF MATRICES STUDIED

$$\begin{array}{c} 1 \\ \begin{array}{cc} \text{I} & \text{II} \\ \end{array} \\ \begin{bmatrix} +31 & +22 \\ +10 & +16 \end{bmatrix} \end{array} \quad \begin{array}{c} 2 \\ \begin{array}{cc} \text{I} & \text{II} \\ \end{array} \\ \begin{bmatrix} +14 & +5 \\ -7 & -1 \end{bmatrix} \end{array} \quad \begin{array}{c} 3 \\ \begin{array}{cc} \text{I} & \text{II} \\ \end{array} \\ \begin{bmatrix} -4 & -13 \\ -25 & -19 \end{bmatrix} \end{array}$$

$$\begin{array}{c} 4 \\ \begin{array}{cc} \text{I} & \text{II} \\ \end{array} \\ \begin{bmatrix} +33 & +17 \\ +12 & +23 \end{bmatrix} \end{array} \quad \begin{array}{c} 5 \\ \begin{array}{cc} \text{I} & \text{II} \\ \end{array} \\ \begin{bmatrix} +11 & -5 \\ -10 & +1 \end{bmatrix} \end{array} \quad \begin{array}{c} 6 \\ \begin{array}{cc} \text{I} & \text{II} \\ \end{array} \\ \begin{bmatrix} -5 & -21 \\ -26 & -15 \end{bmatrix} \end{array}$$

Conditions.—The six payoff matrices used are shown in Table 1. The rational decision rule is the same for all: choose Option I if the probability of the first row event exceeds .4, and choose Option II if this probability is less than .4. The first three matrices differ from each other only by a positive or negative constant added to all entries. The entries are all positive in the first matrix, mixed in the second, and negative in the third. The last three matrices are similarly related. The first three meet the necessary inequalities so that a step function is predicted in response to them. For the second set of three, no prediction from the theory is known. They were run to see whether *S*s appear to react differently to the two classes.

During the 4-day preliminary phase, seven events with probabilities from .2 to .8, spaced in .1 steps, were used. Each of the 42 event-matrix combinations was presented from 9 to 13 times. These data were used to estimate the approximate event probability for which *S*'s choice probability was .5 for each matrix. A range from .1 below to .1 above this point was explored in the experiment. As mentioned above, some mistakes about S_1's ranges were made, but they were rectified shortly after the experiment began. Aside from that, all save one of the other ranges were appropriate.

Fifteen hands were selected in these .2 probability ranges so that the steps between adjacent ones varied from .020 probability at the edges to .008 in the center of the range. Because we did not want *S*s to be aware of differential treatment, events spaced at .05 intervals from .25 to .75 were also presented to all *S*s. Nonetheless, S_4 and S_5 realized that the other *S*s were getting different house hands, and they questioned *E* about this. They were assured that we would distribute the bonus on the basis of the number of points earned relative to the maximum possible.

Each of the 90 event-matrix combinations appeared once every 90 trials for a total of 50 presentations during the experiment. Within a block of 90, their order was randomized. The 36 event-matrix combinations that were added to this basic design to convince *S*s that they were being treated alike were each presented five times. Rows or columns were interchanged randomly on successive presentations of the same event-matrix pair.

RESULTS AND DISCUSSION

Plots of the estimated choice probabilities vs. the theoretical event probabilities for the six payoff matrices of the S_1, S_2, and S_3 are shown in Fig. 1. To save space, the corresponding plots for S_4 and S_5 are not presented because in all cases they are almost perfect simple discontinuities. For S_4 the break occurs at event $p = .4$, which is the rational decision, and for S_5 it occurs at $p = .5$. Thus, the data from S_4 and S_5 are consistent both with the step function hypothesis and with the classical economic assumption of a simple

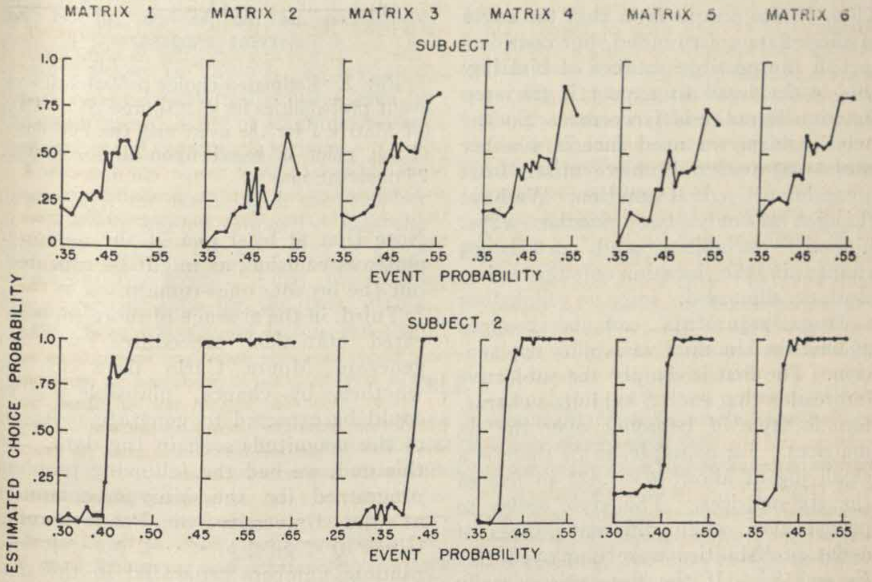


FIG. 1. Estimated choice probabilities vs. event probabilities for S_1, S_2, and S_3 on the six different payoff matrices. (Each point is based upon 50 observations.)

discontinuity, but they are not consistent with the ogival hypothesis (except, of course, in its limiting form). The data for *S*s 1–3 are certainly inconsistent with the assumption of a simple discontinuity, but whether steps or ogives underlie these data is less clear.

Because we do not know how to reduce this question to a well formulated and solved statistical problem, the reader ultimately will have to make his own decision. The following considerations may be helpful.

It is clear that the data exhibit plateaus, some of which are broad (particularly in S_1, Matrices 1–6; S_2, Matrices 3 and 5; and S_3, Matrix 1). There is also no question that the edges of these flats are rounded, but considering all the possible sources of blurring, this is the least to expect if the step function hypothesis is correct. So the only problem we need face is whether these plateaus could have arisen from a continuous, ogival function. We have thought of only two possible ways: binomial variability and an abrupt change in the location of the ogive. Both are discussed.

Three arguments can be brought against the binomial variability explanation. The first is simply the subjective impression that each *S* exhibits a characteristic style of behavior over the six matrices. For example, in S_3 there is a small jog at about $p = .425$ in five of the six matrices. The style seems to persist even when different ranges of event probabilities were employed (see S_2 and S_3). If the flat regions really result only from sampling variability, no internal *S* consistency, such as there seems to be, should exist.

Second, if variability accounts for the plateaus, then they should not stand up very well when the data are divided into two parts. Separating the even and odd observations for each point, we obtain plots such as those in Fig. 2.[3]

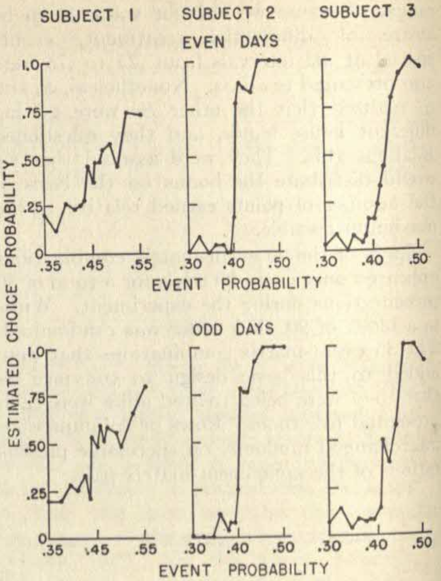


FIG. 2. Estimated choice probabilities vs. event probabilities for S_1, S_2, and S_3 on Payoff Matrix 1 for the even and the odd days. (Each point is based upon approximately 25 observations.)

Note that at least one of the narrower plateaus vanishes, as might be expected, but the broader ones remain.

Third, in the absence of more sophisticated statistical procedures, we can generate Monte Carlo data to see whether, by chance, plausible ogives could be expected to generate reversals of the magnitude seen in the data. To this end, we had the following problem programed for the Univac computer at the University of Pennsylvania.[4] Given $0 < p_1 < p_2 < \cdots < p_{15} < 1$, use random numbers generated in the machine to obtain N (in our case, 50) independent binomial observations for each p_i. Let N_i be the number of successes. For each lag k, $k = 1, 2, \ldots, 8$,

---

[3] As it is impractical to present all 36 of these curves, and as there seems to be no particularly rational way to choose a few from among them, we simply have presented the first matrix for each *S*. The same choice is made in Fig. 3.

[4] We wish to thank Donald Norman and A. Martin Wall for preparing the program and carrying out the runs and we would also like to express our appreciation to the Computer Center of the University of Pennsylvania for its cooperation.

TABLE 2

PROPERTIES OF THE LOGISTIC FUNCTIONS ($P = 1/1+e^{a\sigma}$) USED TO GENERATE THE MONTE CARLO DATA

Ogive Number	σ for which $P = 1/2$	jnd	σ Range
1	.45	.06	.35–.55
2	.40	.02	.30–.50
3	.40	.02	.35–.55

store the number of differences $N_{i+k} - N_i$ which are zero or negative. Repeat this M times (we used $M = 200$) and read out, for each k, how many cases exhibited 0, 1, . . ., and 15-k reversals, i.e., zero and negative differences.

Three sets of probabilities p_i were selected as follows. Three logistic curves, $1/(1 + e^{a\sigma})$, were chosen to approximate most of the data curves; they have the features shown in Table 2. By letting σ run over the event probabilities used in the experiment, a set of 15 probabilities was generated from each logistic.

For each S and each matrix, we determined for each lag k the number of times the choice proportion was the same or smaller for an event k steps above another event. Then, using the Monte Carlo data for the best fitting of the three ogives, we found the percentage of times that many or more reversals of lag k occurred in the Monte Carlo runs. Thus, for example, S_2 in Matrix 4 exhibited six reversals of lag 2, and the Monte Carlo runs for Ogive 3 showed five cases out of 200 with six or more reversals at lag 2. So the entry for this case is $5.100/200 = 2.5\%$. These estimated probabilities, which are summarized in Table 3, measure how likely it is that reversal structure of our data could have arisen by chance from the ogive. It should be kept in mind that the eight entries per row are not independent.

The patterns for S_2 and S_3 seem unambiguous; it is most unlikely that our data could have arisen from a logistic with a .02 jnd. The pattern is less clear for S_1, but we note that except for Matrix 3 there is always at least one lag for which the observed number of reversals has 5 chances or less in 100 of occurring if the ogive is the true function.

On the basis of these considerations, we doubt that the observed plateaus can have resulted simply from ogives plus sampling variability.

A second way a continuous ogive might give rise to a plateau is for the center of the ogive to be located at one event probability on some trials and at another on the remainder. If the absolute value

TABLE 3

PERCENTAGE OF 200 MONTE CARLO RUNS IN WHICH THE NUMBER OF REVERSALS OF LAG k EQUALED OR EXCEEDED THE NUMBER OBSERVED

S	Matrix	Ogive	1	2	3	4	5	6	7	8
1	1	1	2	97.5	3.5	34.5	*	2	*	*
	2	1	77.5	40	3.5	0	0.5	0	*	0
	3	1	77.5	40	28	34.5	*	*	*	*
	4	1	5	76.5	28	34.5	18.5	2	*	0
	5	1	17.5	40	3.5	6	0.5	*	*	*
	6	1	5	14	28	34.5	18.5	*	*	*
2	1	2	0	0	0	0	0	*	*	*
	4	3	0	2.5	1	0.5	1.5	0	0	2
	5	2	0	0	0	0	*	*	*	*
	6	3	0	2.5	0.5	0.5	1.5	2.5	8	2
3	1	2	14.5	0	0.5	0	0	0	0	*
	3	3	0	2.5	0.5	0.5	0	0	0	2
	4	3	0	2.5	0.5	0.5	1.5	2.5	8	*
	5	3	1.5	2.5	0.5	0.5	1.5	0	0	2
	6	3	0	2.5	0.5	0.5	1.5	2.5	8	*

Note.—The asterisk indicates that in these cases no reversals were observed, hence that number was equaled or exceeded in all Monte Carlo runs.

of the difference in locations is large relative to the breadth of the ogive, then the expected data would exhibit a single intermediate step with rounded edges. With three discrete locations for the ogive, two rounded steps can result. And so on. Such discontinuous shifts of location might, for example, correspond to major shifts in *S*'s opinion about the rational breaking point. If these abrupt changes in location occurred irregularly from day to day during the experiment, we know no way of detecting them; however, if the shifts occurred only once or twice during the experiment, then, for example, the data from the first half of the experiment should differ appreciably from those from the second half. Such plots are shown in Fig. 3; the horizontal features of the plots are substantially the same in the two halves.

Although there is no evidence for the sort of shift that would produce plateaus, this figure does suggest that there may have been systematic changes in the underlying probabilities during the experiment. To check this, the choice proportions for the first and second halves of the data were compared for each *S*-matrix combination by the Wilcoxon matched-pairs signed rank test. The *P* levels are summarized in Table 4, where we see that in over half the cases there are significant trends. The effects are not, however, very large. Their main consequence would be to increase the variance of the data slightly beyond binomial.

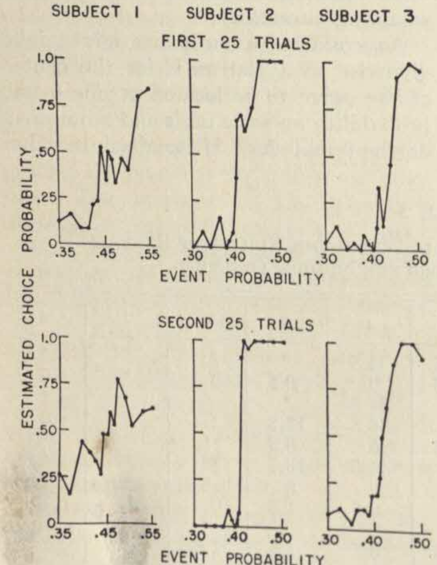


FIG. 3. Estimated choice probabilities vs. event probabilities for S_1, S_2, and S_3 on Payoff Matrix 1 for the first 25 and the second 25 trials. (Each point is based upon 25 observations.)

TABLE 4

P-VALUE COMPARISONS BETWEEN FIRST AND SECOND HALVES OF THE DATA

S	Matrix					
	1	2	3	4	5	6
1	>.05	>.05	.01–.02	>.05	.02	<.01
2	>.05	.01	<.01	>.05	.01	>.05
3	<.01	<.01	.02	.02–.05	.05	.02

Note.—*P* values are based on Wilcoxon matched-pairs signed rank test.

On the basis of the preceding arguments, we are inclined to view the data as supporting the experimental hypothesis that the choice probability for certain pairs of gambles is a step function of the event probability.

It will be recalled that in contrast to matrices, such as 1–3, in which one row dominates the other, no corresponding theoretical prediction about the response probabilities is known for matrices, such as 4–6, in which the two largest payoffs lie on a diagonal. This may just be our ignorance or it may be that the theory is simply noncommittal about these matrices. In any event, the data fail to suggest any behavioral differences. If the predicted steps actually exist, they also appear when neither row dominates the other; if not, then neither class of matrices exhibits steps.

Furthermore, there is no suggestion that the sign of the payoffs differentially affects the qualitative features of the behavior.

Finally, we turn to the other class of data recorded, response times. It is often held that response times, although ill understood, are closely related to response probabilities, and so it is conceivable that they too might exhibit discontinuities. The means and variances of the response times were plotted as functions both of the response probabilities and of the event probabilities for each *S*-matrix pair. These plots are not presented because, so far as we could see, there is no consistent picture. Much of the possible detail is masked by variability, and those trends that do show through differ both within *S*s and within matrices. Probably one cannot expect to make subtle uses of response times in experiments, such as this one, where no strong external time pressures are imposed.

SUMMARY

In this experiment, each *S* chose repeatedly between pairs of gambles. His payoff on each trial was determined by his choice together with the outcome of a chance event which was completely independent of his choice. Our purpose was to test the mathematically derived prediction that, when certain inequalities are satisfied by the payoffs, *S*'s choice probability is a step function of the event probability. As an alternative hypothesis we selected the intuitively plausible one that the function relating these two probabilities is ogival. Although qualitatively these hypotheses contrast sharply, various a priori considerations make it difficult to discriminate between them.

Data were collected from 5 *S*s. Each was given six payoff matrices with 15 events per matrix and 50 observations per event. Two *S*s exhibited almost perfectly discontinuous functions, and so they were (trivially) consistent with the step function hypothesis and only consistent with the ogival hypothesis in its limiting form. The other 3 *S*s produced what can best be called lumpy functions. Arguments were adduced to suggest that neither variability coupled with an underlying ogive nor abrupt shifts in the location of an ogive are sufficient to explain the data. No other ways have been suggested whereby an ogive could generate data having plateaus.

The tentative conclusion is that the data support the step function hypothesis more strongly than the ogival one, and that indirect confirmation is given to the assumptions underlying the prediction.

REFERENCES

ANDERSON, C. D. The constant-ratio rule as a predictor of confusions among visual stimuli of brief exposure duration. Technical Note, 1959, Hearing and Communication Laboratory, Indiana University.

CHIPMAN, J. S. Stochastic choice and subjective probability. In Dorothy Willner (Ed.), *Decisions, values, and groups.* New York: Pergamon, 1960. Pp. 70–95.

CLARKE, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. Acoust. Soc. Amer.*, 1957, **29,** 715–720.

COOMBS, C. H. On the use of inconsistency of preferences in psychological measurement. *J. exp. Psychol.*, 1958, **55,** 1–7.

COOMBS, C. H. Inconsistency of preference as a measure of psychological distance. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories.* New York: Wiley, 1959. Pp. 221–232.

DAVIDSON, D., & MARSCHAK, J. Experimental tests of a stochastic decision theory. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories.* New York: Wiley, 1959. Pp. 233–269.

DAVIDSON, D., SUPPES, P., & SIEGEL, S. *Decision making.* Stanford: Stanford Univer. Press, 1957.

LUCE, R. D. A probabilistic theory of utility. *Econometrica*, 1958, **26,** 193–224.

LUCE, R. D. *Individual choice behavior: A theoretical analysis.* New York: Wiley, 1959.

LUCE, R. D., & RAIFFA, H. *Games and decisions.* New York: Wiley, 1957.

MOSTELLER, F., & NOGEE, P. An experimental measurement of utility. *J. pol. Econ.*, 1951, **59,** 371–404.

PAPANDREOU, A. G. A test of a stochastic theory of choice. *U. Calif. Publ. Econ.*, 1957, **16,** 1–18.

RAMSEY, F. P. *The foundations of mathematics.* New York: Harcourt, Brace, 1931.

VON NEUMANN, J., & MORGENSTERN, O. *Theory of games and economic behavior.* (2nd ed.) Princeton: Princeton Univer. Press, 1947.

(Received October 17, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 50–56

EXTINCTION EFFECTS FOLLOWING NONDIFFERENTIAL REINFORCEMENT OF AN IRRELEVANT STIMULUS[1]

SALLY E. SPERLING[2]

University of Michigan

In any discrimination learning situation, there are other stimuli present on each experimental trial besides the positive and negative discriminanda that *E* manipulates. These are usually called "extraneous" or "irrelevant" stimuli. Several investigators, e.g., Lawrence (1949, 1950), Babb (1956), and Jeeves and North (1956), have placed two experimentally manipulable irrelevant stimuli into discrimination learning problems with rats and investigated learning with respect to these stimuli. The latter two studies, both of which utilized simultaneous stimulus presentation and choice responses, agreed in showing that *S*s who received initial training during which each of the irrelevant stimuli appeared equally often with the positive and negative discriminanda did not learn a test problem, where one of these stimuli is now positive and the other negative, any faster than control *S*s who received initial training without the irrelevant stimuli being present.

In test learning, where the stimuli were presented together on each trial, it is not possible to determine why the experimental groups did not differ from the control groups. This might be either because their previous reinforcement history with respect to the stimuli was ineffective in increasing the probability of response in the presence of these stimuli or because the increased probability of response was equal for the two stimuli since both of them were treated alike in terms of pairings with the positive and negative relevant stimuli and in terms of reinforcement.

That the latter alternative may hold is suggested by the consideration that the irrelevant stimuli may be thought of as being under a 50% partial reinforcement schedule during initial training; i.e., they were each associated with reinforcement and with nonreinforcement half of the time that they were present. If this is sufficient to determine partial reinforcement effects, then the consequences of nondifferential reinforcement of an irrelevant stimulus should be like those observed following 50% partial reinforcement (Lewis, 1960).

The purpose of this study was to examine resistance to extinction in the presence of a previously nondifferentially reinforced irrelevant stimulus, where initial training and extinction were given in a runway under single stimulus presentation to eliminate the difficulties posed by choice responses. If the past reinforcement history with respect to the irrelevant stimulus determines the level of extinction performance, then this group ought not to differ from a group which has had 50% partial reinforcement. Both of these groups ought to show more resistance to extinction than groups which have had either 100% reinforcement or no previous exposure of the stimulus present during extinction.

---

[1] This report is based on a dissertation submitted to the Horace H. Rackhan School of Graduate Studies of the University of Michigan in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The experimental work was supported by a grant from the National Science Foundation, (G-9779), to J. D. Birch, and the writer wishes to thank him for his innumerable helpful suggestions.

[2] Now at the University of California, Riverside.

Method

Subjects.—The *S*s were 55 male hooded rats, aged 90–110 days at the beginning of experimental training, obtained from the colony maintained by the Department of Psychology at the University of Michigan. Four animals were eliminated during the experiment because of illness.

Apparatus.—The apparatus, described in detail by Birch, Ison, and Sperling (1960), consisted of two elevated platforms separated by a partition. The starting platform was 12 in. long and 3 in. wide. The stimulus platform, 18 in. long and 4 in. wide, was divided into two 9-in. sections by a beaverboard partition. A 6½ in. × 4⅛ in. swinging door could be hung from a metal frame directly behind a like-sized aperture cut in the partition; a 6-v. flashlight bulb was mounted on the door, 3¼ in. above the stimulus platform facing the starting platform. An unpainted aluminum food cup could be attached to the terminal end of the stimulus platform.

An infrared light beam obtained from two Worner cells was directed 1 in. over the starting platform, ½ in. from its terminal end. When the light beam was broken by *S*, a 100-w. bulb illuminated the stimulus platform, the flashlight bulb mounted on the swinging door went on, and a GE Standard Electric timer calibrated in .01 sec. was started. When *S* stepped off the starting platform, a normally-off microswitch placed under it was released to stop the clock. With the exception of the floor of the stimulus platform, which was mid-gray floor tile, the entire apparatus was painted flat black.

Stimuli.—Two stimulus platforms longitudinally striped with black and white Mystic tape were used during acquisition training. One of these had stripes 1 in. wide (W); the other had stripes ½ in. wide (N). These two platforms were striped to the swinging door and were covered with longitudinally laid gray tape from the door to their terminal ends. A third platform was completely covered with gray tape. The swinging door was covered with black tape which concealed the mounting of the flashlight bulb. Stimulus platforms could be changed by an arrangement of snap fasteners. During extinction trials the swinging door and its frame were removed. The two stimulus platforms used during extinction were identical to the acquisition striped platforms except that the stripes extended the length of the platforms.

Preliminary training.—There were 14 days of pretraining. The first 6 days were devoted to handling and taming. During the last 8 days *S*s were under 23-hr. deprivation, eating 10 gm. of dog pellets during a 1-hr. feeding period. On the first 5 of these days *S*s were placed on elevated platforms of unpainted pine and allowed to eat from attached food cups. On the last 3 days *S*s received a total of five rewarded trials in the experimental apparatus with an unpainted pine stimulus platform in place. On the two trials of the last day of pretraining, a black painted aluminum door was in place across the stimulus platform.

Experimental groups.—The *S*s were randomly assigned to one of six treatment groups. Five of these received 8 acquisition trials per day for 10 days; the sixth received 8 trials per day for 5 days. For groups wnLD[3] ($N = 9$) and LD ($N = 8$), "light bulb on" (L) was the positive stimulus and "light bulb off" (D) the negative stimulus. For the first of these groups, W and N were each present on half of the positive and half of the negative trials; for the second group, the gray platform was always present. For Groups WNld,a ($N = 10$) and WN ($N = 8$) W was positive and N negative; L and D were each present on half of the positive and half of the negative trials for the former, and D obtained on all trials for the latter. Group wn ($N = 8$) was presented with W and N under the D condition; each of the two striped stimuli was associated with a 50% partial reinforcement schedule. Group WNld,b ($N = 8$) had the same stimuli as Group WNld,a but received only 40 acquisition trials.

All *S*s received 16 extinction trials on the day after their last acquisition day. The light bulb, swinging door, and frame were removed from the apparatus for these trials. Thus the illumination in the stimulus compartment was like that obtaining under the D condition during acquisition. For 4 Group LD *S*s the N platform was presented during extinction. The W platform was present for all others.

---

[3] Upper case letters are used to designate stimuli associated with 100% and 0% reinforcement and lower case letters to designate stimuli associated with 50% reinforcement.

TABLE 1

Acquisition and Extinction Conditions for Experimental Groups

Variable	Groups						
	wnLD $N = 9$	wn $N = 8$	WNld,a $N = 10$	WN $N = 8$	WNld,b $N = 8$	LD $N = 4$	LD $N = 4$
Acquisition stimuli							
Light on	100%[a]	—	50%	—	50%	100%	100%
Light off	0%	—	50%	—	50%	0%	0%
Wide stripe	50%	50%	100%	100%	100%	—	—
Narrow stripe	50%	50%	0%	0%	0%	—	—
Acquisition trials	80	80	80	80	40	80	80
Extinction stimulus	W	W	W	W	W	W	N

[a] Percentage of reinforcement during acquisition trials.

The experimental design is presented in Table 1.

Experimental procedure.—All *S*s received 8 trials per day, of which 4 were rewarded. Each *S* in each of the three complex stimulus groups had two pairings of each of the four possible combinations of stripe and light conditions each day. Gellermann (1933) orders were used for the schedules of presentation by omitting an initial or terminal pair of alternating trials from the 10-trial schedules. A table of random numbers was used to determine the complex stimulus presentations for those groups that had them. Within the limits of differences in group *N*s and stimuli presented, schedules were matched across groups. The light bulb on the swinging door was unscrewed but left in place for those trials or groups where the L condition was not called for.

The *S*s were given experimental trials in groups of 5 to 7, maintaining a constant 15-min. intertrial interval for both acquisition and extinction. During initial training the food cup was baited with two .125-gm. pellets for rewarded trials and *S* was removed after 20 sec., which was sufficient for consumption of the reward. On nonrewarded trials *S* was detained for 20 sec. with no food cup present. The *S*s were under 21-hr. food deprivation during acquisition training and 19¾-hr. deprivation at the beginning of the extinction day.

All statistical analyses were in terms of logarithms of reciprocal latency × 100 (log RS).

Results

Table 2 shows mean log RS on W trials on the last day of acquisition training for Groups wnLD; WNld,a; WNld,b; wn; and WN, and mean log RS over all trials on that day for Group LD, which did not have stripes present during acquisition. An analy-

TABLE 2

Mean Log RS for W Trials on Last Day of Acquisition and for 15 Extinction Trials

Measure	Group					
	wnLD	wn	WNld,a	WN	LD	WNld,b
Acquisition	1.74	2.03	2.03	2.04	1.70[a]	1.66
Extinction	1.36	1.83	1.56	1.12	.88	1.08

[a] Mean log RS over all 8 trials of the last day, since Group LD was not presented with W during acquisition.

sis of variance on the data of the first five groups yielded an *F* of .85 (*df* = 4/38; $P > .20$) indicating that terminal level of responding in the presence of the stimulus to be presented during extinction cannot be considered different among these groups. A second analysis including Group LD yielded an *F* of .88 (*df* = 5/45; $P > .20$), indicating that Group LD's terminal response level may be considered equivalent to that of the other groups.

Since all groups showed a marked increase in mean log RS from first to second trials, and since the first trial was not properly an extinction trial, the first trial was omitted from all subsequent analyses. Two different analyses of the extinction data were made. The first concerns log RS during 15 extinction trials, the second concerns mean decrease in log RS in the presence of W from the last acquisition day to the extinction day.

Group wn ran fastest during extinction, followed in order by Groups WNld,a; wnLD; WN; WNld,b; and LD. All groups showed a decrease in log RS over the extinction trials. Table 3 summarizes an analysis of variance on these trials. For the main effects it can be seen that there were group differences in log RS during extinction and that there was a significant decrease in response speed over groups. The nonsignificant Groups × Trial interaction suggests that the amount of decrease did not differ among groups.

TABLE 3

Analysis of Variance of Log RS for 15 Extinction Trials

Source	*df*	*MS*	*F*
Between *S*s	50		
Groups (G)	5	15.07	3.71*
Error (b)	45	4.06	
Within *S*s	714		
Trials (T)	14	.89	2.62**
T × G	70	.24	.70
Error (w)	630	.34	

* $P < .01$.
** $P < .005$.

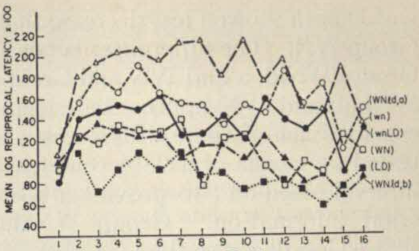


FIG. 1. Mean log RS on 16 extinction trials in the presence of W.

The rejection of the hypothesis of equal log RS among groups allowed individual comparisons to be made. The means for the groups are presented in Table 2. The difference between the means of Groups wnLD and LD yielded a *t* significant at the .05 level. All other differences were tested by 95% confidence intervals based on Tukey's method (Scheffé, 1959) to provide descriptive information. Group wn ran faster than Groups LD; WNld,b; and WN.

For the second analysis of the extinction data, mean log RS for each *S* in each of the experimental groups for its 4 W trials on the last acquisition day was subtracted from its mean log RS over the 15 extinction trials. Group LD was omitted from this analysis since these *S*s did not have W present during acquisition. An analysis of variance performed on the decrease scores indicated that there were differential amounts of decrease among the groups ($F = 3.56$; *df* = 4/38; $P < .05$). Tukey 95% confidence intervals were used to test these differences. Groups wn and wnLD both showed less decrease than Group WN. The differences between Groups WNld, a and WN and Groups WNld,b and WN approached significance. Each of these differences was tested by means of a 98% confidence interval based on *t* to provide a posteriori information. Group WNld,a showed less decrease than Group WN; the other difference was not significant.

DISCUSSION

The following interpretation of the results of this study rests upon the assumption that the primary determinant of extinction performance in the various groups was the reinforcement history associated with W during initial training. The stimulus situation during extinction may be considered to be most closely analogous to W and D, and the reasons for omitting consideration of D should be mentioned. These reasons are: (*a*) During extinction the light bulb, swinging door, and frame were removed from the apparatus. The only way, then, in which the extinction situation resembled the D acquisition condition was in the level of illumination of the stimulus platform. (*b*) Groups wn and WN received all of their acquisition training under the D condition, i.e., they were both 50% reinforced in the presence of D. There is some question as to whether the constant level of illumination would come to exercise any control over *S*s behavior, thus the omission of a lower-case "d" from the designations of these groups. (*c*) Analyses of the past reinforcement history of the WD pattern for the different groups or of the combined reinforcement histories of W and of D separately do not lead to any increased explanatory power with respect to these data than that resulting from the consideration of W alone.

If the past reinforcement history with respect to W is considered to be the primary determinant of extinction performance, then, with the exception of Group WNld,a, the results of this study form a reasonable and fairly simple pattern.

Resistance to extinction following nondifferential reinforcement of an irrelevant stimulus during discrimination training on other stimuli in Group wnLD was greater than the resistance to extinction exhibited by Group LD with no previous reinforcement history with respect to this stimulus and equivalent to the resistance to extinction observed in Group wn following 50% reinforcement.

That the first result is not due to disruption of performance in Group LD as a result of the changed stimulus conditions from acquisition to extinction is suggested by the fact that Groups wn and WN both had the same stimulus change and their extinction performance was clearly different, i.e., the partial reinforcement effect is generalizable to a two-stimulus training situation under a constant 15-min. intertrial interval. In addition, Groups WN and LD had different amounts of stimulus change from acquisition to extinction and there was apparently little difference in their extinction performance.

Group WNld,a, which received 100% reinforcement in the presence of W during acquisition training, should have exhibited less resistance to extinction than Groups wnLD and wn and should not have differed from Group WN. The data, however, indicated that a reverse relationship may hold, since Group WNld,a did not differ from Groups wnLD and wn, the 50% reinforcement groups, and showed less decrease than Group WN, the comparable 100% reinforcement group.

The initial training conditions for Groups wnLD, wn, and WNld,a were alike in one respect not shared by Group WN. All of the first three groups had a 50% partial reinforcement schedule present during acquisition training although this schedule was associated with W and N for Groups wnLD and wn and with L and D for Group WNld,a. It has been demonstrated a number of times that resistance to extinction is greater following 50% reinforcement than following

100% reinforcement under conditions where separate groups have been used for this comparison. To the writer's knowledge, this study is the first in which resistance to extinction has been examined following the concurrent administration of 50% and 100% reinforcement schedules. Under these conditions, response strength, as measured by resistance to extinction, seems to be the same in the presence of a previously 50% and a previously 100% reinforced stimulus, suggesting that resistance to extinction in both cases may be some sort of resultant of the administration of two concurrent schedules.

There were no differences between Groups WNld,a and WNld,b under either analysis of extinction data, indicating that Group WNld,a was still in the range of trials where resistance to extinction is positively related to the number of rewarded trials.

There is no evidence in these data for the operation of the mediational process proposed by Lawrence (1949, 1950) to account for the results of test learning following complex stimulus training. He assumed that relevant stimuli become more distinctive during initial training while irrelevant stimuli either do not acquire or lose distinctiveness. New responses are thus more easily learned to relevant than to irrelevant stimuli during test learning through the mediation of acquired distinctiveness. It is not clear whether responses in the presence of more distinctive stimuli ought to be more or less resistant to extinction than responses in the presence of less distinctive stimuli, but a difference between them is clearly implied. In the present experiment, however, no such difference was observable during extinction in the presence of W between Group WNld,a for which W had been a relevant stimulus, and Group wnLD, for which W had been an irrelevant stimulus.

The absence of a strong discriminative response on the initial problem in both of these groups may provide a complicating factor since it could be argued that the modification of initial orders of distinctiveness does not occur until *S* is performing at a high level of correctness on the initial problem. Alternatively, however, it is possible that an analysis of the training conditions used by Lawrence in terms of reinforcement schedules associated with complex patterns of stimuli and position responses may provide an explanation of his data that makes the appeal to mediational processes unnecessary.

SUMMARY

Five groups of rats were given 80 acquisition trials and one group 40 acquisition trials in an elevated runway under single stimulus presentation, using Wide (W) and Narrow (N) striped stimulus platforms and a light bulb, mounted on a swinging door, that was either on (L) or off (D). All *S*s were given 16 extinction trials in the presence of W with the light bulb and swinging door removed from the apparatus.

The results were: (*a*) Group wnLD, for which W had been a nondifferentially reinforced irrelevant stimulus during discrimination training on L and D showed more resistance to extinction than did Group LD which also received discrimination training on L and D but in the presence of a gray platform. (*b*) Group wn for which W had been a 50% partially reinforced stimulus also showed greater resistance to extinction than Group LD and did not differ from Group wnLD. Group wn was more resistant to extinction than Group WN for which W had been the positive discriminandum, demonstrating that the partial reinforcement effect is generalizable to a two-stimulus training situation.

(*c*) Group WNld,a for which W had also been the positive discriminandum, with L and D present as irrelevant stimuli, did not differ in extinction from Group WNld,b, the 40-trial acquisition group, or from Groups wnLD and wn and was apparently more resistant to extinction than Group WN, the comparable 100% reinforcement group.

Except for these last results, the data are consistent with the interpretation that the primary determinant of extinction performance is the past reinforcement history associated with the stimulus presented during extinction, regardless of the previously relevant or irrelevant status of this stimulus.

Thus nondifferential reinforcement of an irrelevant stimulus may be considered to be equivalent to 50% partial reinforcement.

REFERENCES

BABB, H. Proportional reinforcement of irrelevant stimuli and transfer value. *J. comp. physiol. Psychol.*, 1956, **49,** 586–589.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60,** 36–40.

GELLERMANN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42,** 207–208.

JEEVES, M. A., & NORTH, A. J. Irrelevant or partially correlated stimuli in discrimination learning. *J. exp. Psychol.*, 1956, **52,** 90–94.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39,** 770–785.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40,** 175–189.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57,** 1–28.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

(Received October 19, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 57–61

THE ROLE OF DRIVE (TIME STRESS) IN COMPLEX LEARNING: AN EMPHASIS ON PRELEARNING PHENOMENA [1]

R. ERNEST CLARK [2]

State University of Iowa

The efficacy of Hullian drive theory (Hull, 1943; Spence, 1956) for predicting drive effects in complex learning, though well demonstrated by Spence and his colleagues (e.g., Spence, Farber, & McFann, 1956; Spence, Taylor, & Ketchel, 1956), has been controversial because of some antagonistic evidence (e.g., Hughes, Sprague, & Bendig, 1954), and because of the many assumptions typically surrounding the predictions (Hill, 1957). These assumptions have been of three types: (*a*) theoretical assumptions concerning the interactions between such relevant Hullian variables as habit strength, drive, threshold, and oscillatory inhibition; (*b*) coordinating assumptions required in the application of Hullian drive theory to complex learning, e.g., that "manifest anxiety" (Taylor, 1953) is a valid measure of generalized drive; and (*c*) special assumptions to supplement inadequate empirical data regarding the states of specific task variables, e.g., assumed numbers and relative strengths of expected competing responses.

The present study was designed to eliminate many of the special and coordinating assumptions typically used in predictions of complex learning performance by establishing before learning the actual empirical effects of the drive variable on the numbers and relative strengths of competing responses. Given these empirical data, specifically relevant to the present task and sample of *S*s, it was possible to predict early learning performance without presupposing the correctness of Hullian theory, or its applicability to complex learning.

METHOD

Apparatus[3]

The stimulus consisted of a short vertical bar of light resting on a long horizontal line of light. The bar was viewed by *S* as a movable divider that could be made to cut the line into any two segments. Actually, during the experiment proper, only one position of the bar on the line was used. This position was exactly halfway between center and the left end of the line.

The response unit consisted of a stylus and semicircular response slot, with a "ready" hole for the stylus as the origin of the semicircle. Lining the base of the response slot were ¼-in. wide brass contacts which, when touched with the tip of the stylus, allowed for a determination of the spatial response. A head-rest clamping device permitted *S* to see the stimulus, but not the response slot.

Subjects

Thirty-two volunteers from the elementary psychology classes of the State University of Iowa were assigned to a high-drive group and 32 to a low-drive group. Random assignment of *S*s was restricted by the following selection criteria: (*a*) groups were balanced with regard to three levels of *MA* scale scores (Taylor, 1953) to reduce the possibility of disproportional interaction of "manifest anxiety" with the drive conditions; (*b*) 16 *S*s in each group were men and 16 were women; (*c*) each *S* had a minimum "visual acuity" of 25/20 as measured with the Snellen Eye Chart.

---

[1] This investigation represents part of a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the State University of Iowa (1958).

[2] Now at the Quartermaster Research and Engineering Command, United States Army, Natick, Massachusetts.

[3] The detailed apparatus has been described elsewhere (Gerjuoy & Bechtoldt, 1955).

Preliminary Training

The response required of *S* was to duplicate the divisions seen in the stimulus by pushing the stylus into the response slot to a point which would divide the slot appropriately. Six uncorrected practice trials were given, each trial consisting of a 4-sec. silent period, a 2-sec. "ready" tone, and a 4-sec. stimulus duration. It was emphasized that all of the time available should be used to make an accurate response, but that the response to each stimulus should be completed before the stimulus light went off. The 4-sec. stimulus duration used in this phase of the study allowed *S*s to make rather leisurely responses.

Experiment Proper

The study proper was divided into three sets of 21 trials each, with the reinforcement of selected responses occurring only during the third set of trials. Drive was defined in terms of "time stress," or the time allowed *S* to complete a response. This drive variable has been used successfully before (e.g., Castaneda & Palermo, 1955), and was well suited for the present study since it was easily controlled by varying the duration of the stimulus light.

Trials 1–21 (*initial response hierarchies*).—During this period all *S*s responded 21 times to the 4-sec. stimulus, i.e., under minimal time-stress level. The relative frequencies of occurrence of spatial responses made to the stimulus indicated the structure of the competing response hierarchies for the two groups under a constant, and presumably low, drive level.

Trials 22–42 (*response hierarchies under different drive levels*).—The *S*s continued responding without interruption for another 21 trials while different levels of time stress were induced by the following conditions: Group 1 (reduced time, high time stress): The stimulus duration was shifted from 4 sec. to a duration giving *S* barely enough time to respond. The actual duration used for each *S* was adjusted on the basis of his mean response speed during the first experimental period. Group 2 (control, low time stress): The *S*s of this group continued responding to the 4-sec. stimulus. The same low level of time stress was expected to continue.

The response hierarchies obtained during Trials 22–42 showed the specific influences of varied time stress on competing response frequencies just prior to learning.

Trials 43–63 (*learning*).—During the final 21-trial period, *S*s continued to respond under varied time stress. However, just before the first learning trial, *S*s were informed that all correct responses would now be followed by the sound of a chime, and that responses not followed by the chime were to be considered inaccurate.

During learning, half of the 32 *S*s in each time-stress group learned an easy response and half a difficult response, the relative difficulty of responses being estimated from the previously obtained relative response frequencies. Thus, the easy correct response was selected as *S*'s most probable (dominant) response and the difficult correct response as the very improbable response lying one interquartile range to the right of the dominant response. The interquartile range for each *S* was calculated from the frequency data obtained during Trials 1–21.

RESULTS AND DISCUSSION

Effects of Time Stress on Response Hierarchies

The mean frequencies of various spatial responses elicited by the stimulus before (Trials 1–21) and after (Trials 22–42) time stress was varied are shown in Fig. 1. The center "tick marks" in the abscissas of the figure correspond to each *S*'s dominant, or most frequent, response. The marks to the left and right of center refer to the contact regions in the slot immediately adjacent to the dominant responses.

From these data it was possible to predict the early learning performance of high and low time-stress groups for any selected responses. For example, in the present case of the easy (dominant) correct response, the high time-stress *S*s should have produced more correct responses early in learning than the low time-stress *S*s. This prediction based upon the data in Fig. 1 was reinforced by a

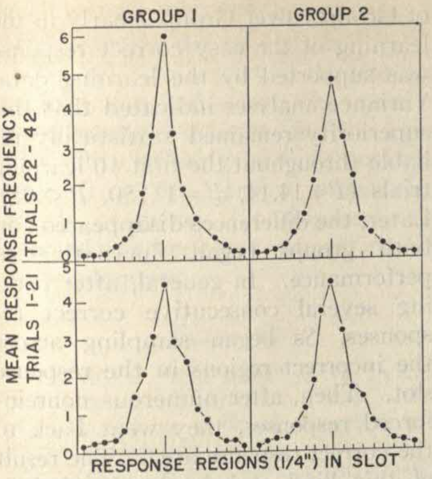


FIG. 1. Hierarchies of competing responses before (Trials 1–21) and after (Trials 22–42) time stress was varied.

statistical analysis of the dominant response frequencies. Variance analyses and subsequent *t* tests indicated that during Trials 1–21 the two groups were equivalent in mean frequencies of dominant responses. When time stress was varied during Trials 22–42, however, a significant increase occurred in the frequency of this response for Group 1 (increased time stress), but not for Group 2 (constant time stress). The means and *SD*s associated with the analyses are shown in Table 1.

Because of the infrequency of the difficult correct response, no comparable statistical analysis was possible. However, since the known prelearning frequency of this response approached zero for both experimental groups, neither group was expected to perform very well early in learning.

TABLE 1

MEANS AND *SD*s OF FREQUENCIES OF DOMINANT RESPONSES BEFORE (TRIALS 1–21) AND AFTER (TRIALS 22–42) TIME STRESS WAS VARIED

Group	Trials 1–21		Trials 22–42	
	Mean	*SD*	Mean	*SD*
1 increased time stress	4.50	2.18	6.06	2.88
2 constant time stress	4.41	2.05	4.47	1.81

Note.—Critical difference based on $P < .05$ between column means is 1.12, and between row means is 1.03.

Before going on to the learning results, it is of interest to examine the effects of varied time stress upon the overall hierarchies of competing responses, including those responses not actually scheduled for reinforcement during learning. If time stress had functioned multiplicatively, changes in response strengths accompanying increased time stress (Trials 22–42) would have been proportional to the original response strengths (Trials 1–21). Generally, strong responses would have increased in frequency and weak responses decreased.

Figure 2 shows the relationship between original response frequencies and those obtained under the varied time-stress conditions. In such a plot of the response variable, a "no-change" reference may be represented by a straight line with slope of one and origin of zero. Had no changes occurred in the respective frequencies of responses from Trials 1–21 to Trials

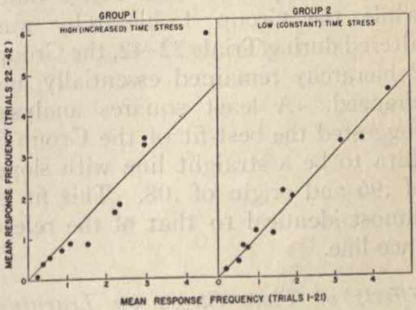


FIG. 2. Relative response frequency during Trials 22–42 as a function of response frequency during Trials 1–21 and time-stress level.

22–42, all of the plotted points would have fallen upon the reference lines.

Considering first the data of Group 1, it is obvious that increased time stress was accompanied by systematic changes in relative response frequencies. These changes appear to be largely consistent with the notion of time stress as a multiplicative drive variable, and consequently, with Hullian drive theory. Stronger responses increased in frequency in proportion to their original relative strengths, and weaker responses decreased in proportion to their original relative weakness. Although the very weak competing responses, initially zero to one in mean frequency, seemed to be unaffected by varied time stress, little confidence may be placed in these data because of the very few scores involved. Even considered collectively, these extremely weak responses represented less than 4/21 of the total response frequency.

It should be made clear at this point that, although the effects of time stress were not contrary to Hullian theory, the present data are not in such form as to support the Hullian multiplicative rule in unequivocal preference to other monotonic functions that may suitably relate drive and habit variables.

It was apparent from Fig. 2 that, while the Group 1 hierarchy was altered during Trials 22–42, the Group 2 hierarchy remained essentially unchanged. A least squares analysis suggested the best-fit of the Group 2 data to be a straight line with slope of .96 and origin of .08. This fit is almost identical to that of the reference line.

Effects of Time Stress on Learning Performance

Learning an easy response.—Figure 3 shows that the predicted superiority of Group 1 over Group 2 early in the learning of the easy correct response was supported by the learning data. Variance analyses indicated that this superiority remained statistically reliable throughout the first 10 learning trials ($F = 14.14$, $df = 1/150$, $P < .05$). Later, the differences disappeared and both groups began fluctuating in performance. In general, after making several consecutive correct responses, *S*s began sampling across the incorrect regions in the response slot. Then, after numerous nonreinforced responses, they went back to the correct response area. The result of this kind of behavior is obvious in the "dips" and "rises" of the functions in Fig. 3.

Learning a difficult response.—As illustrated in Fig. 3, neither group produced many correct responses early in the learning of the difficult response. An analysis of variance indicated no group differences in correct response frequencies throughout the 21 learning trials. There was, however, a significant trials effect ($F = 3.11$,

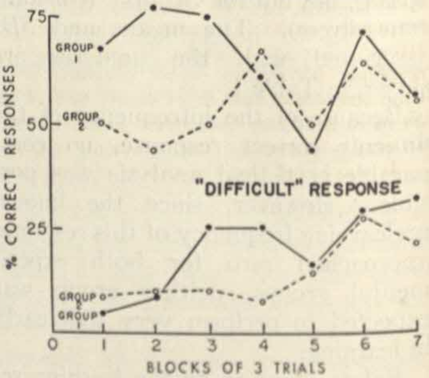


FIG. 3. The percentage of correct responses produced in blocks of three trials by each group in learning the easy and difficult responses. (Group 1 refers to the high time-stress group and Group 2 to the low time-stress group.)

$df = 6/180$, $P < .01$), suggesting that some learning did take place.

The relative brevity of the learning periods resulted from the present experimental interest in the predictability of *early* learning performance from the *known* prelearning effects of drive. To this end, the present methodology proved to be quite adequate.

SUMMARY

The role of "time stress" (drive) in complex learning and performance was investigated by relating empirical changes in prelearning competing response hierarchies to differences in the early learning performance of high and low time-stress groups. The relative strengths of competing responses were provided before learning by relative frequencies of occurrence of spatial responses to a single stimulus in a motor paired-associate task. The response consisted of dividing a slot with a stylus so that the divisions corresponded with those presented in a stimulus line. When given a number of trials, *S*'s responses to the single stimulus were variable enough to yield a hierarchy of competing responses with differing relative frequencies of occurrence.

The 63 trials of the experiment proper were divided into three periods of 21 trials each, reinforcement of selected responses occurring only in the last period. First, an initial response hierarchy was obtained while all *S*s responded under the same low level of time stress. Second, changes in the competing response hierarchies were obtained while time stress was varied. Finally, half of the *S*s in each of two time-stress groups learned an easy response and half a difficult response.

It was found that the effects of time stress (drive) upon prelearning competing response hierarchies were largely compatible with Hullian drive theory. Predictions of early learning performance, based on these known prelearning effects, were supported by the learning data.

REFERENCES

CASTANEDA, A., & PALERMO, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175–179.

GERJUOY, H., & BECHTOLDT, H. P. *Generalization effects in successive discrimination learning as measured by response similarity.* Final Report, March 8, 1955, State University of Iowa, ONR Contract Nonr 1261-(00).

HILL, W. F. Comments on Taylor's "Drive theory and manifest anxiety." *Psychol. Bull.*, 1957, **54**, 490–493.

HUGHES, J. B., SPRAGUE, J. L., & BENDIG, A. W. Anxiety level, response alternation, and performance in serial learning. *J. Psychol.*, 1954, **38**, 421–426.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

SPENCE, K. W., TAYLOR, J. A., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received October 20, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 62–67

THE EFFECTIVENESS OF CONCEPTS AT VARIOUS LEVELS OF AWARENESS

KEITH G. DAVIS AND HARRIE F. HESS[1]
University of Colorado *University of Oregon Medical School*

Throughout the history of psychology there has been continuing interest in phenomena which may be classified under the rubric of "behavior without awareness." It is generally conceded that concepts may be formed without awareness. The recent studies on verbal conditioning (Krasner, 1958) provide contemporary evidence bearing on this point. The issue with which this paper deals is whether or not concepts which operate below awareness can guide behavior efficiently.

One study which has frequently been cited as providing evidence which indicates that unconscious concepts can effectively guide behavior is that of Rees and Israel (1935). Leeper (1951) cites the Rees and Israel experiment as one showing ". . . that concepts may be formed, retained, and used 'unconsciously'." In a recent review, Adams (1957) cites the Rees and Israel experiment as one of the few experiments in one major category providing unequivocal, positive evidence of behavior without awareness. Other authors refer to the Rees and Israel study in a similar manner (Hebb, 1958; Leeper & Madison, 1960; Woodworth & Schlosberg, 1954).

In the Rees and Israel experiment, *S*s who *were not* aware of a system for solving anagrams used the system about as effectively as other *S*s who *were* aware of the nature of the system. This finding that awareness is essentially unrelated to the adaptive use of concepts contradicts contemporary theoretical formulations (Dollard & Miller, 1950; Kelly, 1955; Rogers, 1951; Rotter, 1954), which stress the importance of awareness. This contradiction, taken together with the facts that Rees and Israel used only 10 *S*s, and that their observations of awareness were merely anecdotal, suggested that the pertinent part of the Rees-Israel study be replicated.

Rees and Israel (1935) established a set (or, according to subsequent interpretations, they taught their *S*s a concept) by training them on a series of 15 anagrams, each of which had only one possible solution. The solutions to the training anagrams could be found by rearranging the letters according to a certain letter order. A second series, consisting of 15 multiple-solution anagrams (i.e., anagrams which could be solved by the letter-order pattern or by alternative letter rearrangements), was then presented. The strength of the induced set was measured by counting the number of anagrams in the second series solved according to the letter-order pattern. All of Rees and Israel's experimental *S*s were highly influenced by the "set." They solved 95% of the multiple-solution anagrams according to the letter-order pattern as compared with 47% for 10 control *S*s. The remarkable aspect of the findings was that 6 of the 10 experimental *S*s reported no conscious awareness of having used a pattern to guide their solutions, although 91% of their solutions were according to the pattern. Rees and Israel (1935) concluded that, "The most significant feature of this experiment was its demonstration that the set could operate at such a high level of effectiveness without the subject's being aware of its existence" (p. 15).

The present paper is a report of a replication of the Rees and Israel experiment which includes: (*a*) a larger and more heterogeneous sample, (*b*) a more systematic method of assessing awareness, and (*c*) an additional measure of the effectiveness of the concept. The research hypothesis, derived from contemporary theorizing in the field of personality, was that contrary to the results of the Rees and Israel study, awareness of the pattern or concept would be highly related to effectiveness of the pattern in guiding solutions to anagrams.

---

[1] Now with the Nevada State Department of Health, Las Vegas, Nevada.

METHOD

Subjects.—Seventy-four experimental and 24 control *S*s were studied. The 74 experimental *S*s consisted of four different groups; (*a*) 24 introductory psychology students, (*b*) 18 firemen, (*c*) 10 psychology graduate students, (*d*) 22 high school students from the upper 25% of a senior class. The control group consisted of 24 introductory psychology students.

Procedure.—*S*s were seen individually. The procedure was essentially identical to that used by Rees and Israel (1935). The anagrams were typed in lower-case letters and presented singly by employing a cardboard strip with a small window. The *S* was required to continue on each item until a solution fitting the mechanical requirements was obtained or until 90 sec. had elapsed. At the end of 90 sec., *S* was required to leave the item and go on to the following one. The *E* also timed each solution.

Experimental design.—The design of the experiment was the same as that used by Rees and Israel (1935, p. 14). All *S*s were required to solve 10 practice anagrams with haphazard letter orders before beginning the experiment proper. Following the practice series, all *S*s were required to solve 30 anagrams. For the experimental group, the 30 anagrams consisted of two series of 15 anagrams each: the training series and the test series. The 15 anagrams of the training series were single-solution anagrams with letters of the words rearranged in a 54123 pattern (SCENE is the anagram ENSCE). The test series consisted of 15 multiple-solution anagrams soluble both by the letter-order pattern, and by one or more alternative arrangements.

The 30 anagrams for the control group were made from the same words as those used with the experimental group. The 15 multiple-solution anagrams were presented to the control group in form identical to the form of presentation to the experimental group, but they were interspersed with the single-solution anagrams. The 15 single-solution anagrams had letters arranged in haphazard order. The actual list of anagrams used and the order of presentation can be found in Rees and Israel (1935, pp. 24–25).

The reader will note that the method of presentation of the "test series" (i.e., the 15 multiple-solution anagrams) was different for the experimental and control groups. Consequently, differences between experimental and control groups on the response measures cannot confidently be attributed to the effects of the training series. This defect in the original experimental design is inconsequential to the purpose of the present paper, which is concerned only with differences between *S*s within the experimental (trained) groups.

Measures of the influence of training.—Two measures were used to determine the effectiveness of the letter-order pattern in guiding the solutions of the experimental groups on the test anagrams. The first measure was the same as used by Rees and Israel (1935), viz., the number of multiple-solution anagrams solved according to the letter-order pattern by each *S*. The second measure, response latency, was the mean time required for each *S* to solve the multiple-solution anagrams.

The measurement of awareness.—Rees and Israel (1935) used a postexperimental interview to obtain a dichotomous measure of the extent to which *S*s were aware of the letter-order pattern. To overcome the disadvantages of the interview and to provide a graduated, rather than a dichotomous measure of awareness, the present experiment utilized a 4-point awareness rating scale similar to that devised by Matarazzo, Saslow, and Pareis (1960). In this scale, awareness ratings were based upon *S*s' written replies to three open-ended sentences mimeographed on two sheets of paper, presented separately. The sentences were: (Sheet 1) "The purpose of this experiment is . . .," and "My evidence for this is . . .," and (Sheet 2) "The method I used was . . . ." Judges independently rated the written responses of *S*s on a 4-point rating scale according to predetermined criteria of awareness patterned after those used by Matarazzo, Saslow, and Pareis (1960).

The reliability of the awareness scale was determined by having four judges independently rate the written replies of the 18 *S*s in the firemen group. An index of order association (Senders, 1958) was calculated between each pair of raters. The indices were found to be .96, .98, .98, .98, 1.00, and 1.00 ($P < .001$ for an index of .96), indicating a satisfactory reliability of rating. The rater with the highest agreement with other raters completed the awareness ratings for the remainder of *S*s. All ratings were made without knowledge of the performance of the *S*s on the anagram series.

RESULTS

Effectiveness of the training procedure.—Table 1 presents the mean

TABLE 1

EFFECTIVENESS OF TRAINING PROCEDURE

Group	*N*	Mean Number of Prescribed Solutions		Mean Latency in Seconds	
		Mean	*SD*	Mean	*SD*
Replication study:					
Controls (introductory psychology students)	24	9.08	2.30	22.76	8.47
Experimental groups:					
Introductory psychology students	24	14.38	1.55	5.25	7.24
Graduate students	10	14.00	1.63	4.34	3.27
Firemen	18	10.61	3.42	21.55	19.18
High school students	22	12.82	2.75	11.02	11.84
Combined groups	74	12.95	2.86	10.81	13.72
Original Rees and Israel study:					
Experimental group	10	14.19[a]		Not Reported	
Controls	10	7.10[a]		Not Reported	

[a] Estimated from data presented by Rees and Israel (1935).

scores of the experimental and control groups on the two response measures. It can be seen that the combined experimental (trained) group showed (*a*) a larger number of anagrams in the test series solved according to the prescribed pattern, and (*b*) shorter latencies than the control group ($P < .001$, Mann-Whitney *U* Test). Table 1 also presents estimates of the mean number of prescribed solutions for the experimental and control groups of the original Rees and Israel (1935) study. It can be seen that these scores for *S*s of the original experiment are generally similar to those for the comparable experimental groups (the graduate student groups and the introductory psychology student groups) and the control group of the present study. Thus, it may be concluded that the results of the

TABLE 2

NUMBER AND PERCENTAGE OF *S*S IN EACH AWARENESS CATEGORY

Awareness Rating	Individual Experimental Groups								Combined Experimental Groups		Control Group (Introductory Psychology Students)		Rees & Israel Experimental Group[a]	
	High School Students		Introductory Psychology Students		Graduate Students		Firemen							
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Level 1 (unaware)	5	23	1	4	0	0	5	28	11	15.5	17	71	6	60
Level 2	1	4.5	0	0	0	0	4	22	5	6.5	0	0	3	30
Level 3	4	18	6	25	3	30	3	17	16	21.5	7	29	0	0
Level 4 (fully aware)	12	54.5	17	71	7	70	6	33	42	56.5	0	0	1	10
Total	22	100	24	100	10	100	18	100	74	100	24	100	10	100

[a] Estimated from data presented by Rees and Israel (1935).

replication study are similar to those of the original study with respect to the effect of the training procedure on roughly comparable subject samples.

Degree of awareness of the letter-order pattern.—Table 2 presents the distributions of awareness ratings of both experimental and control *S*s of the present experiment, as well as estimated awareness ratings of Rees and Israel's (1935) original experimental group. The awareness ratings of the Rees and Israel group were made by the present authors from brief descriptions included in the original Rees and Israel report, and must be considered with caution.

In Table 2 it can be seen that 84.5% of the combined experimental *S*s of the present study attained some degree of awareness of the letter-order pattern, whereas only 40% of Rees and Israel's experimental *S*s gained any awareness. Presumably part of this difference is attributable to the more sensitive method of assessing awareness which was used in the present study.

The relation between awareness and the response measures.—Since the two college student experimental groups (introductory psychology and graduate students) were so uniformly high in awareness (Table 2) it was impossible to examine the relationship in these groups between awareness and response measures. However, the awareness ratings and response measures for firemen and high school students display a wider variation, thus permitting a test of the experimental hypothesis. The statistic used was the index of order association. All indices for these two experimental groups were highly significant ($P < .001$) as shown in Table 3. A definite, positive relationship is evident for the firemen and high school groups between awareness and number of prescribed

TABLE 3

RELATIONSHIP BETWEEN AWARENESS RATINGS AND RESPONSE MEASURES

Group	N	Index of Order Association		z Score	
		Prescribed Solutions	Latency	Prescribed Solutions	Latency
Firemen	18	.73	−.40	7.96*	4.12*
High school students	22	.77	−.83	8.80*	9.20*
Combined experimental groups	74	.77	−.70	28.7*	24.8*

* $P < .001$.

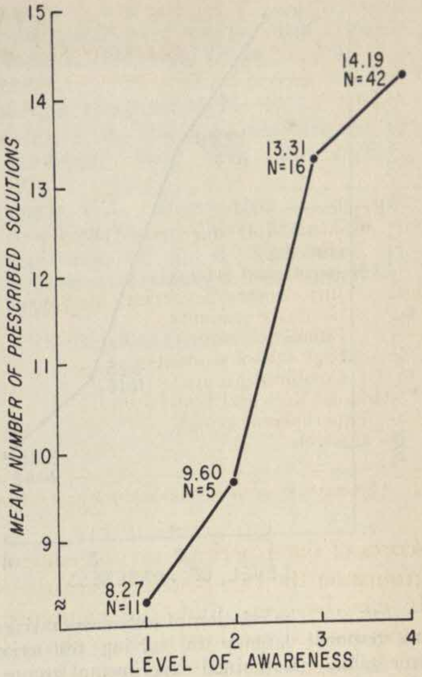


FIG. 1. The relation of awareness ratings to number of test-series anagrams solved according to the letter-order pattern. (Combined experimental groups, $N = 74$.)

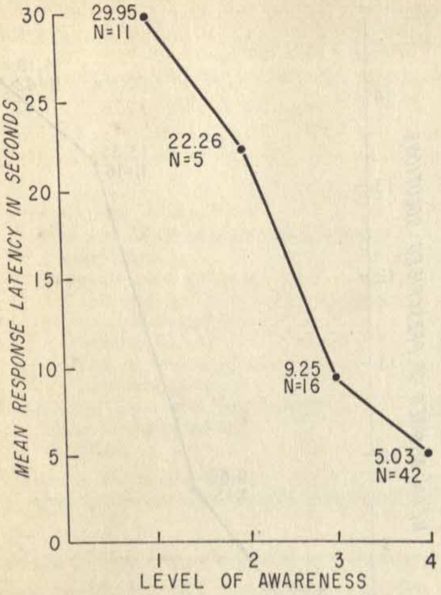


FIG. 2. The relation of awareness ratings to response latency in solving test-series anagrams. (Combined experimental groups, $N = 74$.)

solutions on the test series, as hypothesized. Also, as expected, there is a significant inverse relationship between awareness ratings and response latency.

The last row in Table 3 indicates that the data, when grouped, show the same results as did the separate analysis of the firemen and high school *S*s. Awareness ratings are highly related to both measures of response efficiency ($P < .001$). Figures 1 and 2 depict these relationships graphically.

DISCUSSION

The results of the present replication differ markedly from those of the original study with respect to the relationship between awareness of the letter-order pattern and the extent of use of that pattern in solving anagrams. Among the Rees and Israel (1935) experimental *S*s, essentially no relation was found between awareness of the pattern and the influence of that pattern on solutions to the anagrams; in the experimental groups of the present study, the relation between awareness and usage of the pattern was high and significant. In the broad range of *S*s sampled in the present study, not 1 *S* produced results which suggest an effective, unconscious concept. Thus, the present replication strongly contradicts the Rees and Israel (1935) findings, and in so doing finds itself in agreement with contemporary theoretical formulations concerning the relationship between awareness and adaptiveness of behavior (Dollard & Miller, 1950; Kelly, 1955; Rogers, 1951; Rotter, 1954). The differences found in the relationship of awareness and effective behavior are thought to be due to: (*a*) the improved measure of awareness used in the present study, and (*b*) the larger and more heterogeneous sample used.

SUMMARY

This experiment was a replication of the oft-cited study by Rees and Israel (1935) which reported that 6 out of 10 (60%) of experimental *S*s developed and used highly effective concepts that were below the level of awareness. In the replication more *S*s were used than in the original experiment, and a more systematic method of assessing awareness was employed. None of the *S*s in the replication study developed efficient concepts at a level below awareness. Instead a close relation between awareness and anagram solving efficiency was noted, contrary to one of the main implications of the Rees and Israel study.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, 54, 383–405.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

HEBB, D. O. *A textbook of psychology.* Philadelphia: Saunders, 1958.

KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

Krasner, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

Leeper, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 730–757.

Leeper, R., & Madison, P. *Toward understanding human personalities.* New York: Appleton-Century-Crofts, 1960.

Matarazzo, J. D., Saslow, G., & Pareis, E. N. Verbal conditioning of two response classes: Some methodological considerations. *J. abnorm. soc. Psychol.*, 1960, **61**, 190–206.

Rees, H. J., & Israel, H. E. Establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**(6, Whole No. 210).

Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.

Rotter, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

Senders, V. L. *Measurement and statistics.* New York: Oxford Univer. Press, 1958.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Wiley, 1954.

(Received October 21, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 68–73

SIZE-CONSTANCY JUDGMENTS AND PERCEPTUAL COMPROMISE

V. R. CARLSON[1]

National Institute of Mental Health

A generally accepted formula for expressing experimental size-constancy results is the Brunswik (or Thouless) ratio, which reflects the extent to which an obtained size-judgment departs from retinal angular size relative to the difference between a correct retinal value and a correct physical-size value (Brunswik, 1956, p. 21). Although the formula itself allows for the occurrence of numerical ratios outside this range, the derivation and intent of the formula are clearly based on the conception of two essentially limiting values between which perceived size normally falls (Brunswik, 1933, pp. 387–411; Thouless, 1931, pp. 343–344).

Empirically this range seems to cover size-judgments more or less adequately. In the nonreduction situation instructions to judge on the basis of apparent size typically produce average values which fall short of perfect size-constancy ("underconstancy"); analytic, or projective, instructions lead to average values considerably closer to, but still larger than, the lower limit corresponding to a correct retinal value (Brunswik, 1956, pp. 67–69; Gilinsky, 1955; Holaday, 1933; Joynson, 1949; Singer, 1952). These results lend credence to a general view that perceived-size is a compromise between retinal image-size and physical object-size (Brunswik, 1956, pp. 47–48; Osgood, 1953, p. 276; Woodworth & Schlosberg, 1954, p. 483).

Instructions to judge in terms of objective, physical-size, however, are likely to result in ratios somewhat greater than 1.00 ("overconstancy"). The average discrepancy in this direction is not usually as great as the deviation from 1.00 in the opposite direction obtained in a projective match, and overconstancy has not seemed to be of equal theoretical importance. It has become increasingly clear, however, that overconstancy is not a rare phenomenon which occurs only under very special circumstances. Its existence is as firm as the existence of underconstancy and cannot be dismissed simply as error (Carlson, 1960; Gilinsky, 1955; Holway & Boring, 1941; Jenkin, 1957; Smith, 1953).

The general hypothesis proposed here is that the principle underlying the Brunswik and Thouless ratios is incorrect. Rather than size-constancy representing a limiting "pole of intention" for size-judgment (Brunswik, 1956, pp. 98–99), it represents the basic functional relationship between object-size and perceived-size. It is the natural neutral point from which size-judgment starts, and substantial deviations from this value can be produced in *either* direction by variations in experimental conditions.

In particular the present experiment is concerned with determining whether overconstancy will occur to as great an extent with appropriate instructions as does underconstancy with analytic instructions. It was previously hypothesized that the perspective attitude operates to produce

---

[1] The author is greatly indebted to Eugene Tassone for assistance in the collection and analysis of the data.

overconstancy with objective-size instructions (Carlson, 1960). It follows that if the perspective attitude is made explicit in instructions to *S*, the effect should be enhanced. At the same time there is no basis either in the immediate visual stimulation or in logic for such perspective instructions to cause any effect different from objective-size instructions. If a difference does occur, therefore, it must be due to an attitudinal bias on the part of the observer.

METHOD

Subjects.—The *S*s were 32 junior college students, 23 men and 9 women, who were paid for serving as *S*s. They were divided into four equal groups, treated exactly the same except for the order in which they received the various instructions. One *S* could not understand the perspective-size instruction (see below) and was not used in the experiment.

Procedures.—Each *S* was tested individually on six different occasions with several days to 1 week intervening between successive sessions. Group 1 received objective-size instructions on Days 1 and 2, perspective-size instructions on Days 3 and 4, and apparent-size instructions on Day 5. Group 2 followed the same schedule but with the order of the objective-size and perspective-size instructions reversed. Groups 3 and 4 were comparable to Groups 1 and 2, respectively, except that the apparent-size instruction occurred on Day 1 instead of Day 5. The apparent-size judgments were included for comparison and correlational reasons secondary to the main purpose of the experiment.

All *S*s received instructions to match for projective-size on Day 6. This task was reserved until last, because it specifically defines the required size-estimate as something different from a size-constancy judgment and would very likely influence the results in the other conditions, none of which, in and of themselves, require any deviation from size-constancy. Finally, following the projective-size matches, *S* was asked to make two distance estimates. In the first, he indicated the subjective location of the midpoint between himself and the apparatus of the standard triangle. In the second, he adjusted the slider on a 40-cm. model to duplicate the distance of the variable triangle relative to the distance of the standard triangle.

Groups 1 and 2 were tested concurrently, as were Groups 3 and 4. The testing of Groups 3 and 4, however, was not started until Groups 1 and 2 had been completed.

In each condition *S* was required to adjust the size of a variable triangle relative to the following altitudes of a standard triangle: 76, 117, and 159 mm., and four values on either side of 117 mm. spaced in steps of 3.2 mm. All values of the standard were presented randomly twice each, requiring a total of 22 trials at each session.

Apparatus.—The triangles were white and isosceles with an altitude-to-base ratio of 1.39. Each was mounted on an apparatus providing continuous variation in size of the triangle against a 30 × 36 in. black felt background. The standard triangle was located at a distance of 40 ft. from *S*, the variable triangle at 10 ft. and at an angle of 20° to the line of sight to the standard. Each apparatus was placed on a table, so that both triangles were approximately at *S*'s eye level. The *S*'s control knob was mechanically coupled to the variable triangle.

Testing was carried out in a well-lighted room similar to a 10 × 50 ft. corridor except for the presence of numerous miscellaneous objects.

Instructions.—The *S* was initially told that the standard triangle would be changed in size from trial to trial at random, that the variable would always be set for him at a randomly determined starting point, and that he could adjust the variable back and forth until he was reasonably satisfied with the match but not to worry unduly over any single setting. Natural, free binocular regard was specified, and *S* was not allowed to utilize devices such as closing one eye or gauging visual extent by sighting between the fingers. The specific instructions were:

> *Apparent-size.*—Adjust the variable so that it looks equal to the standard in apparent visual size. It may also be equal in actual physical size at that point, or it may not—we are not concerned about that. Try to adjust the variable so that it appears equal to you visually, whether you think it is equal in actual size or not.
>
> *Objective-size.*—Adjust the variable so that it is, as best you can judge, equal in actual physical size to the standard, so that if you were to measure them both with a ruler they would measure the same size. They may also look equal to you in apparent visual size, or they may not—we are

> not concerned about that. Try to adjust the variable so that you think it is the same actual size as the standard whether it appears equal to you visually or not.[2]
>
> *Perspective-size.*—As you look down a road or a set of railroad tracks, the edges of the road or the rails appear to come together in the distance. It is that degree of apparent convergence that we want you to try to duplicate here. Set the near triangle so that if the two triangles were lined up, lines extending from the corners of the near triangle, through the corresponding corners of the far triangle, would seem to come together in the distance. Make this setting so that the apparent convergence of these lines would be the same as the apparent convergence of railroad tracks or the edges of a sidewalk or road which extends away from you.

Some repetition and paraphrasing of this instruction was usually required, such as substituting "way in which the lines or railroad tracks seem to come together" for "apparent convergence." The *E* did not provide any sketches or other illustrations.

Projective-size.—The *S* was instructed to adjust the variable triangle so that its subtended angle would be equal to the angle subtended by the standard triangle. The *E* demonstrated that the image of a small extent close to the eye can encompass the visual image of a more distant, large object and illustrated the geometry of the problem by pencil sketches. As much explanation as necessary was given to ensure that *S* clearly understood the result he was to try to achieve in terms of equality of angular, projective-size. Any statement, however, to the effect that object-size should be proportional to distance was carefully avoided.

RESULTS

The ratio of the physical size of *S*'s setting of the near, variable triangle to the physical size of the far, standard triangle was computed for each trial. The average of these ratios for the three main sizes of the standard (76, 117, and 159 mm.) constituted *S*'s score for a given day. A ratio greater than 1.00 indicates "overconstancy," while a ratio less than 1.00 indicates "underconstancy." This ratio is simply that between the sizes of the variable and standard triangles, not a Brunswik or Thouless ratio.

Scores for perspective-size and for objective-size were each obtained on 2 different days in order to evaluate the reliability of the perspective-size judgments. This reliability ($r_{\text{II}} = .72$) was reasonably comparable to that ($r_{\text{II}} = .85$) for the objective-size judgments (both values based on the entire group of 31 *S*s). The means for the first and second scores were similar in both cases, and the two scores within each condition have been averaged together for all further analyses.

There was no demonstrable effect of order of receiving instructions with respect to whether *S*s received the perspective size instruction prior to the objective-size instruction or vice versa. The *S*s (Groups 3 and 4) who received the apparent-size instruction first, however, clearly manifested a lower mean value for that instruction than those (Groups 1 and 2) who received it later in the schedule (Table 1). This was an unexpected result and requires further comment (see below). But no other significant within-conditions difference occurred among these groups, so the several groups have been combined into a single group of 31 *S*s for all comparisons not involving the apparent-size scores. The statistical significance test employed in each case was a *t* test between correlated means.

The mean objective-size ratio was greater than 1.00, and the mean perspective-size ratio was, in turn, greater than the objective-size value. The projective-size mean was less than 1.00. These differences were all highly significant ($P < .001$). The deviations of the perspective and projective ratios from 1.00 were compared, and the mean difference between the two distributions was clearly insignificant. On the average, *S*s in the perspective situation produced a size ratio as much greater than 1.00 as they did less than 1.00 in the projective situation. The r between the perspective and objective judgments was .47 ($P < .01$); that between the perspective and projective judgments, $-.36$ ($P < .05$). These questions were the primary ones of the experiment, and the results are unequivocal with respect to them.

---

[2] The rationale for these apparent-size and objective-size instructions has been given previously (Carlson, 1960).

A number of previous groups of *S*s have been tested in the present stimulus situation with the same apparent-size instruction. The mean ratios for these groups have been near 1.00, and none has been significantly less than 1.00. Thus, it is the apparent-size mean for Groups 3 and 4 (Table 1) which is anomalous and not that of Groups 1 and 2.

In trying to determine the reason for this discrepancy, it was ascertained that many of the *S*s in Groups 3 and 4 were personally acquainted with individuals who had been *S*s in Groups 1 and 2. It is possible that they had been alerted to the analytic, projective approach to begin with, and it is unfortunate that all the *S*s were not tested with the other instructions prior to testing any with the projective instruction. If this possibility is correct, there should be a correlation between the apparent-size and projective-size judgments for Groups 3 and 4. This correlation was, in fact, .69 ($P < .01$). Whatever the reason for it, these *S*s tended to respond to the apparent-size and projective-size instructions similarly. The corresponding correlation coefficient for Groups 1 and 2 was .37 ($P > .05$).

The remaining intercorrelations among the various conditions, including the distance judgments, were also computed. None reached a statistically significant level, except that the two kinds of distance judgment (midpoint between *S* and standard, and ratio of variable to standard distances) were significantly related to each other ($r = -.64$, $P < .01$).

The nine values of the standard triangle spaced closely around the middle value were utilized to obtain a measure of within-*S*s precision, or sensitivity to relatively small differences in the standard. The rank-order correlation coefficient between these nine values and *S*'s corresponding responses to them was calculated for each *S* for each condition. These coefficients averaged across *S*s were .74, .80, .76, and .76 for the perspec-

TABLE 1

AVERAGE RATIO OF *S*'S SETTING OF NEAR, VARIABLE TRIANGLE TO PRESENTED SIZE OF FAR, STANDARD TRIANGLE

Groups[a]	Measure	Size Instruction			
		Perspective	Objective	Apparent	Projective
1 and 2 ($N = 16$)	Mean *SD*	1.43 0.29	1.19 0.17	1.02 0.18	0.62 0.20
3 and 4 ($N = 15$)	Mean *SD*	1.41 0.23	1.22 0.19	0.79 0.18	0.52 0.17
	p[b]	*ns*	*ns*	<0.01	>0.10

[a] Groups 1 and 2 received the apparent-size instruction after having completed the perspective and objective judgments; Groups 3 and 4 received the apparent-size instruction first.

[b] Probability associated with the t value for the difference between uncorrelated means.

tive, objective, apparent, and projective conditions, respectively. They are all significantly different from zero but do not vary significantly among themselves. Hence there is no reason to suspect that *S*s performed the task any more or less carefully under one instruction than another.

DISCUSSION

The assumption made in the present theoretical argument is that the perspective attitude (Gibson, 1952) is a crucial determinant in specifying a psychologically adequate definition of size in the size-constancy situation. The normal, adult *S*, having lived in a culture which widely uses perspective as a pictorial and communicative representation of object distance, comes to the laboratory already equipped with this attitude. Given this response bias, overconstancy results if *S* is oriented toward judgment in terms of objective-size; underconstancy, if he is oriented toward apparent-size, provided the latter is understood to be something different from objective-size (Carlson, 1960, 1962). The apparent-size instruction used here was specifically designed to eliminate this attitudinal bias, resulting in perfect constancy, at least on the average.

By hypothesis, the perspective instruction devised for the present experiment makes more explicit the bias presumed to be already operating in the objective-size situation. Logically this instruction defines a size match which is no different from an objectively equal match. Since there is a clear increase in the overestimation of size, however, some subjective attitude or assumption must be involved. If it is the same subjective factor present in the objective-size condition, the two sets of judgments should be correlated. The correlation obtained ($r = .47$) was clearly significant.

Similarly, in this view, instructions to match in terms of projective-size make the bias more explicit in the opposite direction. At the same time, this instruction defines a geometrical result which is clearly different from size-constancy and which could be accomplished simply by producing a size ratio which is the same as the distance ratio involved. Some *S*s do seem to try to go about the task this rationally, while others do not. Thus, there is an additional factor which precludes a high correlation with the perspective-size judgments. The lower magnitude of the correlation ($r = -.36$) is therefore reasonable.

Whether this account of the operation of instructions is correct or not, however, the notion that perceived-size is a compromise between physical object-size and retinal image-size does not encompass the occurrence of overconstancy. The further demonstration that overconstancy can occur to as great an extent as underconstancy supports the proposition that size-constancy represents the natural midpoint, rather than one extreme, of the range of experimentally-obtained relationships between perceived-size and object-size. If this hypothesis should prove to have sufficient generality, it suggests two further important implications. The first is that the primary requirement for an adequate theoretical account of size perception is a formulation which specifies invariant perceived-size with varying distance, not increasing or decreasing perceived-size with increasing distance. The second implication is that purported effects of motivational or other personological factors on perceptual constancy would be more adequately formulated as judgmental response biases than as effects upon the perceptual process itself.

SUMMARY

A "perspective-size" instruction designed to make explicit the response bias postulated to be present in size-constancy judgments of objective-size clearly produced a greater degree of overestimation than objective-size instructions. The difference from a physically correct size-match was as great as that obtained in the opposite direction when *S*s judged in terms of projective-size. Perspective-size was positively correlated with objective-size, negatively correlated with

projective-size. It is argued that size-constancy represents the basic functional relationship between object-size and perceived-size, in preference to the view that perceived-size is a compromise between object-size and angular retinal-image size.

REFERENCES

BRUNSWIK, E. Die Zugänglichkeit von Gegenständen für die Wahrnehmung und deren quantitative Bestimmung. *Arch. ges. Psychol.*, 1933, **88**, 377–418.

BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

CARLSON, V. R. Overestimation in size-constancy judgments. *Amer. J. Psychol.*, 1960, **73**, 199–213.

CARLSON, V. R. Underestimation in size-constancy judgments. *Amer. J. Psychol.*, 1962, in press.

GIBSON, J. J. The visual field and the visual world: A reply to Professor Boring. *Psychol. Rev.*, 1952, **59**, 149–151.

GILINSKY, A. S. The effect of attitude upon the perception of size. *Amer. J. Psychol.*, 1955, **68**, 173–192.

HOLADAY, B. E. Die Grössenkonstanz der Sehdinge bei Variation der inneren und äusseren Wahrnehmungsbedingungen. *Arch. ges. Psychol.*, 1933, **88**, 419–486.

HOLWAY, A. H., & BORING, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21–37.

JENKIN, N. Effects of varied distance on short range size judgments. *J. exp. Psychol.*, 1957, **54**, 327–331.

JOYNSON, R. B. The problem of size and distance. *Quart. J. exp. Psychol.*, 1949, **1**, 119–135.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

SINGER, J. L. Personal and environmental determinants of perception in a size constancy experiment. *J. exp. Psychol.*, 1952, **43**, 420–427.

SMITH, W. M. A methodological study of size-distance perception. *J. Psychol.*, 1953, **35**, 143–153.

THOULESS, R. H. Phenomenal regression to the real object. *Brit. J. Psychol.*, 1931, **21**, 339–359.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received October 24, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 74–83

THE OPERATION OF SET IN A VISUAL SEARCH TASK[1]

HARRIET FOSTER

*Engineering Psychology Laboratory, Institute of Science and Technology,
University of Michigan*

The purpose of the present study was to investigate the effects of "mental sets" on the speed of detection or identification of critical components of a visual display. Different types of sets with varying levels of specificity were investigated in order to determine their effects on visual search time and in the hope of gaining some understanding concerning the nature of set. A set is treated here as synonomous with an hypothesis, an expectancy of certain objects, features, or events. Set implies a restriction of the number of possible events for which one is "prepared." A particular set may be described or defined by the unifying concept or the category of objects, characteristics, or events for which the individual is prepared.

Two types of set were investigated in the present study: (*a*) functional sets, in which the categories were based on the meaning or significance of category members without regard to physical appearance, and (*b*) structural sets, in which the categories were based on the physical appearance of members. These two types of set are comparable to the functional and formal categories described by Bruner, Goodnow, and Austin (1956).

Bruner (1957) has suggested that set is a gating or filtering process whereby an individual is sensitive to certain stimulus features or certain events and relatively insensitive to others. One might expect that the efficiency of any particular set would depend upon the number of "gates" included in the set. It was hypothesized in the present study that functional categories require a "gate" for each category member, but that the gates for structural categories consist of the defining attributes or the observable common features of the category members.

With a functional set, then, the time required to recognize an individual category member is expected to increase as the number of category members increases, or as the category becomes broader, more general, or more inclusive. There are a good number of studies which support this hypothesis. Postman and Bruner (1949) and Freeman and Engler (1955) compared a double set for "food" and "color" words with a single set for color words. The double set, of course, contained more alternatives than the single set. Postman and Bruner found a higher mean recognition threshold for the multiple set; but Freeman and Engler found no significant difference between the multiple and single set. Rees and Israel (1935) compared two single sets with a double set which contained fewer alternatives than either single set and found the double set to be more effective than either of the single sets. The prediction concerning functional sets is also consistent with the numerous studies relating reaction time or duration threshold to stimulus or response uncertainty (e.g., Bricker, 1955; Hyman, 1953; Hyman & Hake, 1954; Krulee, Podell, & Ronco, 1954; Krulee & Weisz, 1955) and studies showing decreases in accuracy with increasing numbers of alternatives (e.g., Bruner, Miller, & Zimmerman, 1955; Long, Henneman, & Reid, 1953; Long & Lee, 1953a, 1953b; Long, Reid, & Garvey, 1954.)

---

[1] This work was conducted by Project MICHIGAN under Department of the Army Contract DA-36-039 SC-78801, administered by the United States Army Signal Corps. Particular thanks are due Sally Wurtz for her preparation of the word lists used in this study. The writer is also indebted to William L. Hays for his comments and suggestions concerning the statistical analysis of the data. Subjects in the pilot study were run by Sally Wurtz and in the main experiment by John T. Burns.

Structural sets, on the other hand, are expected to result in increasing recognition times as the number of common attributes of category members increases. In terms of the generality of the category this means an increase in recognition time as the category becomes less inclusive or more specific. Binder's (1955) statistical model for visual recognition makes a similar prediction, but it is difficult to find much supporting empirical evidence. Some indirect support can be found in Chapman's (1932) results showing a loss on attributes for which *S* is not set, and in the results of Lawrence and Laberge (1956) which showed a slight (but not significant) advantage for a set for only one dimension over a set for all dimensions of a stimulus. More direct support can be found in a study by Archer (1954) in which *S*s were required to classify each of a series of visual patterns on four dimensions. The time required to classify the patterns was found to increase with the number of relevant stimulus dimensions and at the same time did not change significantly with different amounts of irrelevant information. On the other hand, evidence summarized by Miller's (1956) "chunking" analysis suggests that a complex stimulus may be responded to as a unit. In this case, no differences in recognition time would be expected as a result of varying numbers of defining attributes of sets or categories.

In addition to the set variables two recognition tasks were studied. Subjects were asked to detect or locate the category member or they were required to identify or specify the particular category member which was present. It was hypothesized that with functional sets, detection would be equivalent to identification. If the gates of a functional set represent individual category members, then the matching of incoming information with one gate is equivalent to the identification of the object. On the other hand, it was expected that detection would be faster than identification when a structural set was used, and that the difference would be greatest for the broadest set or the smallest number of defining attributes. Here, it is supposed, the gate ignores the unique features of one member—the features which must be used to distinguish or identify the particular category member.

METHOD

Stimulus materials.—Printed words were used as the stimulus materials. The functional characteristics (meanings) of printed words are well-known to *S*s as are their structural characteristics (taken here to be the letters which make up the words). The relative inclusiveness of functional categories was varied by selecting general categories and subcategories of the more general categories, whereas the inclusiveness of structural categories varied with the number of letters included in the set. Printed words may represent the most extreme case of a lack of correspondence between structure and function and because of this they may be the best medium for studying the ability of *S*s to respond to functional characteristics of visual targets.

Six functional categories, each with two subcategories, were chosen and lists of four-, five-, and six-letter words falling within these categories were compiled. The Thorndike-Lorge (1944) general word count was used to equate word familiarity of two sets of three lists each (Lists A-1, B-1, and C-1, and Lists A-2, B-2, and C-2). A list compiled by Cohen, Bousfield, and Whitmarsh (1957) was used in an attempt to equate the degree of association of the words with their respective categories, but there were not sufficient empirical data for complete control of this variable.

Each word was presented on a different slide along with eight unrelated words of equal length in a 3 × 3 display, marked off by grid lines into nine cells. The unrelated or background words varied in familiarity (Thorndike-Lorge count), with approximately the same distribution of background word frequency on each slide. No one word appeared on more than one slide. The positions of critical and background words were randomly assigned to the nine locations with the restriction that critical words within any one category list would occur in each of nine positions at least once and not more than twice. All letters of words were capitalized.

Response measures.—Because the number of response alternatives will vary with the broadness of the category or between detection and identification, an attempt was made to control response choice time by allowing *S*s but one response—that of pressing a hand switch held in the left hand. The time measured was the time required by *S* to indicate that he "knew" the correct answer but not the time required to make the correct response. It was hoped that in this way differences between experimental groups would be associated primarily with "perceptual" rather than with response speed. A timer was started when the slide came on; the timer stopped and the slide went off when *S* pressed his hand button.

The *S*s were divided into two response groups—the Detection group (Group D) and the Identification group (Group I). Detection *S*s were instructed to indicate the location of the correct word. As soon as *S* knew the location he pressed his hand switch, which turned off the timer and slide. He then indicated the location of the correct word by pressing with the right hand one of nine buttons on a panel in front of him. The nine buttons were arranged in a 3 × 3 matrix corresponding to the arrangement of words in the slide.

Identification *S*s were instructed to tell *E* what the correct word on the slide was. These *S*s pressed the hand switch as soon as they knew what the word was, and then said the word aloud to *E* after the timer had stopped.

Procedure.—The experimental design is summarized in Table 1. Groups D and I were each subdivided into three set groups. The different lists or categories were presented in the same order for all subgroups. One order of presentation within lists or categories was used for half of the group and the reverse order for the other half of the group. The order in which the different sets were adopted

TABLE 1

EXPERIMENTAL DESIGN

Group	Category or List								
	Functional Categories						Structural Categories		
	A-1 (Animals) 15 Words	A-2 (Transportation) 14 Words	B-1 (First Names) 15 Words	B-2 (Plants) 14 Words	C-1 (Geographical) 15 Words	C-2 (Food) 14 Words	1 9 Words	2 9 Words	3 9 Words
Detection 1_d	Specific Set	Specific Set	Broad Set	Broad Set	Narrow Set	Narrow Set	Very Narrow Set	Broad Set	Narrow Set
Detection 2_d	Narrow Set	Narrow Set	Specific Set	Specific Set	Broad Set	Broad Set	Narrow Set	Very Narrow Set	Broad Set
Detection 3_d	Broad Set	Broad Set	Narrow Set	Narrow Set	Specific Set	Specific Set	Broad Set	Narrow Set	Very Narrow Set
Identification 1_i	Double Set	Double Set	Broad Set	Broad Set	Narrow Set	Narrow Set	Very Narrow Set	Broad Set	Narrow Set
Identification 2_i	Narrow Set	Narrow Set	Double Set	Double Set	Broad Set	Broad Set	Narrow Set	Very Narrow Set	Broad Set
Identification 3_i	Broad Set	Broad Set	Narrow Set	Narrow Set	Double Set	Double Set	Broad Set	Narrow Set	Very Narrow Set

Note.—A warm-up and posttest task involving four-digit numbers preceded and followed the main experiment (see text).

and the particular set which was used for any one list or category differed for the different subgroups. Three types of set were used for the functional categories in Group D: (*a*) a specific set in which *S* was told the exact word for which he was to look, *E* pronouncing and spelling each word; (*b*) a narrow set in which *S* was told the subcategory to which the word belonged (e.g., in Category A-1, members of the animal kingdom, two subcategories were used—four-legged animals and birds); and (*c*) a broad set which included the two subcategories as well as other subcategories (e.g., the animal category included fish as well as four-legged animals and birds). Each narrow or broad category was defined for *S* and then illustrated by three (narrow categories) or four (broad categories) examples. The sets for structural categories were either broad (the first letter specified), narrow (the first two letters specified), or very narrow (the first three letters specified).

The sets prescribed for the subgroups of Group I were the same as Group D, with the exception of the specific set. Because Group I was instructed to specify the exact word which appeared, a specific set instruction would have been meaningless; *S* would have been told exactly what he was to say. In place of a specific set, subjects of Group I were given a double set. They were instructed to look for words which fell into either of the two subcategories. For example, for Category A-1, *S*s were told to look for a word which was *either* a four-legged animal *or* a bird. The sets for structural categories were the same as Group D.

Subgroups differed only in the sets which were adopted. The critical words, their positions on the slides, and their background words were identical for all groups, with the following exceptions: When *S*s were given a broad functional category an additional five slides were introduced to insure that *S*s included in their sets additional subcategories. When *S*s used a narrow or a very narrow structural set, two (narrow set) or one (very narrow set) extra slides were introduced which contained a word in the background which began with the same letter or same two letters as the critical word but which did not contain all of the critical letters. The responses to these extra slides are not included in the data analysis.

The sets assigned for the different categories and groups are indicated in Table 1.

All groups received 27 trials on a warm-up task and 27 trials on a posttest. The stimuli on these two tasks were four-digit numbers arranged in a 3 × 3 matrix as in the case of the words. On each trial *S* was told the first two digits of the correct number. The *S*s of Group D were required to indicate the location of the correct number after they had turned the slide off by means of the hand switch. The *S*s of Group I said the entire number aloud after turning off the slide with the hand switch. The major purpose of the warm-up trials was to reduce variability due to practice in the main experimental trials. The posttest was introduced in an attempt to get some measure of practice or fatigue effects from the beginning to the end of the experimental session. The two tests may also give some indication of individual differences in reaction times within and between groups.

Subjects.—The *S*s were 72 University of Michigan students who had answered an advertisement in the college paper. They were paid according to the amount of time spent and the quality of their performance. As much as was possible, *S*s were assigned successively to the different subgroups, so that 1 *S* was run in each group before a second *S* was run in any group. The subgroups were composed of 12 *S*s each, 6 men and 6 women, with the exception of Group I-2 and Group I-3 and Group D-1, each of which contained 7 men and 5 women.

Apparatus.—The *S* was seated at a console which contained a rear projection screen 8 in. high and 11.5 in. wide. A sloping table projected 10.5 in. from below the projection screen. In the center of this table was the panel containing the 3 × 3 matrix of buttons which *S*s in Group D used to indicate the location of the critical word or number. This response panel was covered by a black box when *S*s in Group I were run. The *S*s held in their left hands the switch which was easily pressed with one finger. They sat in a swivel chair, the legs of which were in a fixed position. The height of the chair was adjusted for different *S*s so that their eyes would be approximately in line with the center of the projection screen. The *S*s wore earphones during the experiment, through which white noise was fed except when *E* was giving instructions. Instructions were also given through the earphones.

The *E*'s panel on the opposite side of the console consisted of a 35-mm. slide projector with magazine loading, a time clock calibrated in .01 sec., a row of nine numbered lights which indicated which of the nine location buttons had been pressed by *S*, and a microphone and foot switch which were used for giving instructions and turning off the white noise.

TABLE 2

MEANS OF LOG MEDIAN RESPONSE TIMES FOR CORRECT RESPONSES ON FUNCTIONAL CATEGORIES

Set	List												Set Mean
	A-1		A-2		B-1		B-2		C-1		C-2		
	Group	Mean	Group	Mean	Group	Mean	Group	Mean	Group	Mean	Group	Mean	
Detection Task													
Broad	3	2.269	3	2.306	1	2.272	1	2.317	2	2.369	2	2.302	2.306
Narrow	2	2.286	2	2.325	3	2.216	3	2.258	1	2.331	1	2.283	2.283
Specific	1	2.114	1	2.073	2	2.116	2	2.094	3	2.087	3	2.086	2.095
List mean		2.223		2.235		2.201		2.223		2.262		2.224	
Identification Task													
Broad	3	2.290	3	2.320	1	2.223	1	2.286	2	2.367	2	2.276	2.293
Narrow	2	2.277	2	2.343	3	2.202	3	2.301	1	2.316	1	2.320	2.278
Double	1	2.289	1	2.302	2	2.232	2	2.286	3	2.327	3	2.285	2.295
List mean		2.285		2.322		2.219		2.291		2.337		2.264	

Note.—Group means for detection task were: 1 = 2.232, 2 = 2.212, 3 = 2.204; group means for identification task were: 1 = 2.274, 2 = 2.305, 3 = 2.287.

Instructions.—Before the warm-up trials *S* was given instructions either to locate or to identify the correct stimulus. He was told that his score would depend upon how quickly he made the correct response and that he would get no credit for incorrect responses. Before each category, *S* was instructed as to what the category was. In the case of the specific set or the structural sets he was told the word or letter(s) preceding the presentation of each slide. For the specific set the word was pronounced and spelled.

RESULTS

Median times for correct responses, in .01 sec., were calculated for each *S* for each of the nine category lists. In order to test the normality of the distribution of medians, each median was transformed to a standard score for the particular task, list, and set in which it occurred. The standard scores were then combined into one distribution and tested for normality by means of a χ^2 analysis. Median times resulted in a distribution of scores significantly different from normal ($P = .01$). Log transformations of median scores, tested in the same manner, did not differ significantly from a normal distribution ($P > .10$). All analyses were therefore made with the log transformed scores.

Functional categories.—Mean scores for the functional categories are shown in Table 2. Warm-up and posttest scores appear in Table 3. Basically the experimental design consists of four latin squares, one for each task and one for each type of category set. However, a latin square analysis for the functional sets would require that pairs of lists be combined. Because the two lists of each pair differed in word frequencies the lists

TABLE 3

Means of Log Median Response Times on Warm-up and Posttest

Group	Warm-up		Posttest	
	Mean	*SD*	Mean	*SD*
D-1	2.191	.067	2.182	.064
D-2	2.210	.051	2.174	.060
D-3	2.179	.048	2.177	.065
I-1	2.199	.034	2.189	.038
I-2	2.208	.069	2.206	.064
I-3	2.202	.082	2.186	.055

were not combined, but, instead, a cross-over analysis was used (Cochran & Cox, 1957). In this way it was possible in one analysis to treat the six lists separately. Table 4 presents the results of the analysis for Groups D and I. Variations in broadness of set result in significantly different response times in Group D but not in Group I. It should be recalled, however, that Group D, used a specific set when Group I used a double set. The significance of the results in Group D appears to be due primarily to the difference between the specific set and the two broader sets. The mean across the six lists is higher for the broad set than for the narrow set, but the difference is small. The same direction of difference between broad and narrow sets is found in the performance of Group I. In both cases there are reversals in the direction of the difference on two of the individual lists. Group differences may be partially responsible for these reversals.

Although the double set results in a slower response time on the average than the narrow set, the difference is not significant, nor is the difference consistent for all lists.

The number of errors made with functional sets was small but showed the same order of difficulty of sets. The number of possible errors with all words, lists, and *S*s combined is 1044 for each task and set. Detection *S*s made 17 errors with the broad set, 9 errors with the narrow set, and 4 errors with the specific set. Identification *S*s made 19 errors with the broad set, 15 errors with the narrow set, and 23 errors with the double set.

Examination of the size of the differences between sets on different lists suggests a possible interaction effect between sets and lists. The experiment was not designed to test interaction effects, and because of the lack of independence between cells a really adequate test can not be made. An attempt was made, however, to discover the presence of an interaction. Each of the three groups on each task was subdivided into six groups of 2 *S*s each. Each pair of *S*s contained 1 *S* from each of the two slide presentation orders. Otherwise the *S*s were randomly paired. Each

TABLE 4

Effects of Variations in Broadness of Set

Source	Functional Categories			Structural Categories		
	df	Group D *F*	Group I *F*	*df*	Group D *F*	Group I *F*
Set	2	127.37**	1.64	2	3.45	35.26*
Lists	5	1.90	20.79**	2	.56	4.12
Groups	2	4.91*	5.28*	2	4.50	59.19*
Error	8	(.000631)[a]	(.000266)[a]	2	(.000327)[a]	(.0000215)[a]

[a] Error mean squares.
* $P < .05$.
** $P < .01$.

pair was randomly assigned to one of the six list and set combinations of its group. Such a procedure, of course, reduces the reliability of the results to a marked degree, as the data from any 1 *S* is being used for only one of the six lists. However, it is possible in this way to obtain independence of the cells and examine the data for interaction effects. The analysis of variance involves a double classification, with three variations of set and six different category lists. Groups D and I were analyzed separately due to the difference in set for the specific and double categories. The analysis for Group D again showed a significant difference between sets, but no significant list variation was found for Group I. The results of the analyses failed to show significant interactions between lists and sets. However, because the number of observations included in the analyses was so greatly reduced, the possibility of an interaction effect remains.

Task differences were not expected on the functional categories. Four separate analyses of variance, required by the lack of independence between cells, were made to test differences between Groups D and I on three lists. Separate analyses were made for the broad and narrow sets and the two sets of matched-frequency word lists. The *F* values for the four analyses are not, of course, independent of each other, as the same *S*s are being compared in each. The four *F* values for the task variables were all less than 1. The *F* values for lists were significant at the .01 level in three of the analyses but were not significant for Lists A-2, B-2, and C-2 when the broad set was used. There were no significant interactions between lists and tasks.

Structural categories.—Mean log times for structural categories are shown in Table 5. The latin square analyses for Groups D and I are presented in Table 4. Group I shows an increasing response time as the structural category becomes more

TABLE 5

MEANS OF LOG MEDIAN RESPONSE TIMES FOR CORRECT RESPONSES ON STRUCTURAL CATEGORIES

Set	List						Set Mean
	1		2		3		
	Group	Mean	Group	Mean	Group	Mean	
Detection Task							
Broad	3	2.127	1	2.155	2	2.162	2.148
Narrow	2	2.193	3	2.150	1	2.217	2.187
Very narrow	1	2.164	2	2.194	3	2.151	2.170
List mean		2.161		2.166		2.177	
Identification Task							
Broad	3	2.120	1	2.159	2	2.167	2.149
Narrow	2	2.184	3	2.148	1	2.175	2.169
Very narrow	1	2.175	2	2.202	3	2.163	2.180
List mean		2.160		2.170		2.168	

Note.—Group means for detection task were: 1 =2.179, 2 =2.183, 3 =2.143; for identification task: 1 =2.170, 2 = 2.184, 3 = 2.144.

specific, or as the number of defining attributes increases. The effect of variation in broadness of set is significant ($P=.05$). Although Group D tends to respond more slowly with the narrow set than with the broad set, times for the very narrow set are somewhat faster than those for the narrow set. The differences in broadness of set are not significant for Group D. The total number of errors for the different sets and tasks is too small to have any significance. The largest number of errors is a total of 3 out of a possible 324 for the narrow set in each of the recognition tasks.

It was predicted that Group I would be slower than Group D when structural sets were used. However, analyses of variance, computed separately for each of the three sets, yielded task *F*s all less than 1. Interactions between sets and lists were tested in the same manner as for the functional categories, but because there were only three lists instead of six the task groups were subdivided into three groups of 4 *S*s each. And because all three set variations were comparable in Groups D and I, it was possible to use a triple classification of tasks, lists, and broadness of set. None of the *F*s was significant. The reader is reminded, however, that the number of observations was reduced to one-third of the original number in order to obtain independent cell scores.

Comparison of structural and functional categories.—The present experiment was designed to measure the effects of variations in the broadness of set within structural and functional categories and not to provide a direct comparison between structural and functional categories. The words used for the two types of category differed, and no direct comparison can be made. However, a small number of *S*s was run in a pilot study in which each of 31 words was tested under each of three sets—broad functional, narrow functional, and broad structural. Functional sets resulted in times significantly longer than times for structural sets.

DISCUSSION

It was hypothesized that the efficiency of set in a visual search task would be a function of the number of "gates" which were required by the set. The gates for functional categories were expected to be the individual category members, leading to increasing recognition times as the category became broader or more inclusive. Variation in structural set was expected to vary with the number of defining attributes of the category, and response times were expected to increase as the category became more specific. The results of this experiment tend to support the hypotheses, but to a limited degree.

Differences between broad and narrow functional sets were small. Although it is not possible to measure the difference in information value between the broad and narrow sets, it seems safe to assume that the difference is much smaller than the difference in information value be-between narrow and specific sets. A more adequate test of the broad-narrow set difference might be made by selecting a narrow category with fewer members, thus increasing the difference in the number of alternatives of broad and narrow categories. For example, a narrow category under the general category of animals might be "mammals" or "members of the cat family." One would expect the effect of differences in set to be more pronounced in such a comparison.

It is possible that the failure to find significant differences for functional categories in the identification task is a function of the nature of the task, rather than of the small differences in broadness of set. Such a possibility seems remote, however, when one considers the lack of difference between detection and identification performance.

The results of this study give little indication of the direction of difference between broad and double sets. If the number of category members is the only determinant, then the broad set should be less efficient. However, it is possible that the disjunctive character of the double set introduces additional difficulties not found with a single set.

It was originally hypothesized that the use of functional categories would result in separate gates for each category member, whereas the different gates for structural categories would be the defining attributes of the categories. The hypothesis concerning structural categories however, is contrary to the fairly large body of literature summarized in Miller's (1956) chunking analysis. Furthermore, the predictions concerning functional and structural sets are inconsistent with each other for very specific categories. A completely specified word if it is considered as a one-member functional category requires the use of only one gate, whereas if it is treated as a specific structural category it requires a gate for each letter. Although the detection times for the very narrow structural set are not significantly shorter than those for the narrow structural set, the results suggest that a switching point may be reached in structural sets. If either type of gate can be used, and the results of this study indicate that they can, then the most efficient procedure with structural sets would be to gate in terms of attributes until the number of attributes becomes larger than the number of category members (supposing that the two types of gates are in a sense equivalent) at which point the gating should then be done in terms of category members.

No significant difference was found in this study between the detection and identification tasks, either with functional or structural categories. The identification required in this task was much more simple than is often the case. All of the cues necessary for accurate identification were present, easy to discriminate, and unambiguous. Perhaps a more difficult identification would have resulted in differences, although there is nothing in the data presented here to indicate that this would be so.

SUMMARY

The study was designed to determine the effects of varying sets on the location and identification of words in a field of unrelated words. Sets were defined in terms of either functional or structural categories with varying degrees of specificity.

In general, response times to functional categories decreased as the categories became more specific, or as the number of category members decreased. Differences between specific word definitions and the broader category definitions were pronounced. Differences between broad general categories and their narrower subcategories were small, but in the predicted direction. Words defined by structural categories tended to be identified more rapidly when the category contained the smallest number of defining attributes. Some ambiguity was found with greater numbers of defining attributes.

The results are interpreted in terms of a "gating" mechanism, where the "gates" for functional categories are the category members and the gates for structural categories are the defining attributes of the category.

REFERENCES

ARCHER, E. J. Identification of visual patterns as a function of informational load. *J. exp. Psychol.*, 1954, **48,** 313–317.

BINDER, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62,** 119–129.

BRICKER, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49,** 73–81.

BRUNER, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64,** 123–152.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

BRUNER, J. S., MILLER, G. A., & ZIMMERMAN, C. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, **49,** 187–192.

CHAPMAN, D. W. Relative effects of determinate and indeterminate Aufgaben. *Amer. J. Psychol.*, 1932, **44,** 163–174.

COCHRAN, W. G., & COX, G. M. *Experimental designs.* (2nd ed.) New York: Wiley, 1957.

COHEN, B. H., BOUSFIELD, W. A., & WHITmarsh, G. A. *Cultural norms for verbal* *items in 43 categories.* Technical Report No. 22, August 1957, University of Connecticut, ONR Contract Nonr-631(00).

Freeman, J. T., & Engler, J. Perceptual recognition thresholds as a function of multiple and single set and frequency of usage of the stimulus material. *Percept. mot. Skills*, 1955, **5**, 149–154.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Hyman, R., & Hake, H. W. Form recognition as a function of the number of forms which can be presented for recognition. *USAF WADC tech. Rep.*, May 1954, 54–164.

Krulee, G. K., Podell, J. E., & Ronco, P. G. Effect of number of alternatives and set on the visual discrimination of numerals. *J. exp. Psychol.*, 1954, **48**, 75–80.

Krulee, G. K., & Weisz, A. Studies in the visual discrimination of multiple-unit displays. *J. exp. Psychol.*, 1955, **50**, 316–324.

Lawrence, D. H., & Laberge, D. L. Relationship between recognition accuracy and order of reporting dimensions. *J. exp. Psychol.*, 1956, **51**, 12–18.

Long, E. R., Henneman, R. H., & Reid, L. S. Theoretical considerations and exploratory investigation of "set" as response restriction. *USAF WADC tech. Rep.*, December 1953, 53–311.

Long, E. R., & Lee, W. A. The influence of specific stimulus cuing on location responses. *USAF WADC tech. Rep.*, December 1953, 53–314. (a)

Long, E. R., & Lee, W. A. The role of spatial cuing as a response-limiter for location responses. *USAF WADC tech. Rep.*, December 1953, 53–312. (b)

Long, E. R., Reid, L. S., & Garvey, W. D. The role of stimulus ambiguity and degree of response restriction in the recognition of distorted letter patterns. *USAF WADC tech. Rep.*, May 1954, 54–147.

Miller, G. A. The magical number seven, plus or minus two: Some limitations on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Postman, L., & Bruner, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369–377.

Rees, J. J., & Israel, H. E. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**(6, Whole No. 210).

Thorndike, E. L., & Lorge, L. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received October 24, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 84–90

ANTICIPATORY TIMING OF CONTINUOUS AND DISCRETE RESPONSES[1]

JACK A. ADAMS AND LYLE R. CREAMER

University of Illinois

Helson (1949, p. 496) appears to have been the first to observe that response to a repetitive input signal in continuous visual tracking tasks has timing errors that are less than would be expected on the basis of visual-motor reaction time (RT). With practice, *S* learns to anticipate regular changes in the input and he achieves a higher performance level than when anticipation is absent and response awaits the visual detection of a change, thus producing tracking error proportional to RT. This type of learned anticipation in paced tasks without preview is what Poulton (1952a, 1952b, 1957) calls *perceptual anticipation.* Adams and Xhignesse (1960) make the additional distinction of *beneficial anticipation*, and define it in a measure of learned response timing that is associated with high skill and is distinct from premature anticipations that can produce performance poorer than if *S* merely had responded after the occurrence of a signal change as in classical RT. Bartlett (1951), Vince (1953, 1955), and Adams (1961) provide further discussions of anticipation in skilled performance.

There is nothing but unverified hypotheses on the underlying mechanisms of perceptual anticipation. Vince (1953) discusses the cognitive basis of these learned anticipatory mechanisms. Adams (1961) discusses as one possibility an S-R hypothesis where an anticipatory overt tracking response is determined by the action of a mediating response and its response-produced cues, which is an explanation similar to that advanced for the effects of verbal pretraining (Arnoult, 1957; Goss, 1955; Vanderplas, 1958). If this hypothesis deserves serious consideration, we should be able to demonstrate that pretraining in anticipation with a nontracking response should show positive transfer to measures of perceptual anticipation for the tracking response in the criterion task. Our experiments reported here were designed to go beyond a demonstration of transfer by further asking whether mediating responses are central-verbal or peripheral-motor responses. Of course the mediation view could be wrong, and anticipation could depend on time-varying concomitants of the criterion motor response itself, and not pretrainable as mediational views would require. If so, we would expect no differences for our various pretraining activities.

EXPERIMENT I

Method

The basic task was one-dimensional continuous pursuit tracking, with the criterion measure being *S*'s anticipation of directional change in a regular input signal which approximated a sine wave. Experimental groups were given special pretraining in anticipation of directional change and then were transferred to the whole task. Comparisons on the criterion response were with a control group having only whole-task practice.

Apparatus.—The signal to be tracked was a cam-generated red line on moving white paper which appeared in a narrow horizontal window below eye level measuring 7 mm. by 30 cm. Wave frequency was 34 cycles per min., and the scale of the signal gave it a lateral sweep of 20 cm. The standing *S* responded with a horizontal handle having a radius of 16.6 cm., and the motor movement was a back and forth arc at about the level of an ordinary table. There was a mechanical linkage between the control and a blue pen writing on the moving paper, and tracking error at any moment was the discrepancy between the red and blue lines. The *E* used two Standard Electric timers for controlling the duration of trials and intertrial rests.

---

[1] This research was supported by the United States Air Force under Contract No. AF 49(638)-371, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

For pretraining in anticipation, auxiliary response apparatus included a throat microphone for verbal responding, a button of 2 cm. diameter mounted next to the control hand for discrete motor responses, and a small crank with a radius of 7.5 cm. for continuous responding. The particular application of each of these will be explained below, but for all of these conditions we followed the procedure of deceiving *S* to think that his pretraining responses were being recorded. This ego-involving approach was used to minimize desultory responding which did not involve active attempts to anticipate directional changes of the signal. Toward this end, each of these pretraining response devices had obvious wires leading from them and out a hole in the wall, and *S* was told that his performance was being measured. These methods were patterned after those of Adams (1955) where similar deceptions were used successfully in guaranteeing *S*'s active participation.

Procedure.—Each trial was 60 sec. in duration and intertrial rest was 30 sec. The three experimental groups had 12 pretraining trials, a 12-min. rest, and 10 whole-task criterion trials. The control group (WT) had only the 10 whole-task trials.

The following is a description of the three different pretraining activities:

Group V (verbal).—The *S* responded by saying the word "Change" into the throat microphone each time the target line reversed direction. He was instructed to anticipate the change so his verbal response coincided with the exact instant of the directional reversal of the signal. The purpose of this group was to see if the mediators were verbal-central and independent of the arm-hand effector used in tracking.

Group CM (continuous motor).—An *S* had a small crank set in a plane perpendicular to the larger control handle of the criterion task. The task was to match the speed and direction change of the input signal, and *S*'s response was a back and forth movement of the small crank over a span of about 180°. Gibbs (1954a, 1954b) has discussed the continuous functions for neural discharge when a voluntary movement is made, and it is plausible that the mediating mechanism for anticipation depends on some aspect of this continuous neural output of the primary motor effector system used in tracking. A pretraining series, where *S* attempts to duplicate the rate and direction of the signal, might prove to be favorable conditions for the acquisition and transfer of anticipatory behavior.

Group DM (discrete motor).—The *S* pressed a button each time the target reversed direction. Like Group V, they were urged to anticipate the directional change so that their press of the button coincided exactly with the change. This was a control for Group CM to cover the possibility that any response with the motor effectors of the transfer tracking task is a sufficient condition for acquiring anticipatory behavior.

Subjects.—The *S*s were 47 university male undergraduates, and they were paid for their participation. The *S*s were randomly assigned to groups. Each group had 12 *S*s except Group DM which had 11. One *S* was discarded because of failure to understand the instructions.

Results

The method of measurement was the same as used by Poulton (1952b) and Adams and Xhignesse (1960). The response time measure was taken as the time in seconds between a directional change of the input signal and a corresponding change in direction of *S*'s response movement as he attempts to follow the reversal. We followed the algebraic convention of a positive sign when response preceded the signal change (lead error) and a negative sign when it followed the change (lag error). A zero value was perfect anticipation. Starting with the third cycle on a trial, measures were taken for the next 60 responses to directional change of the signal.

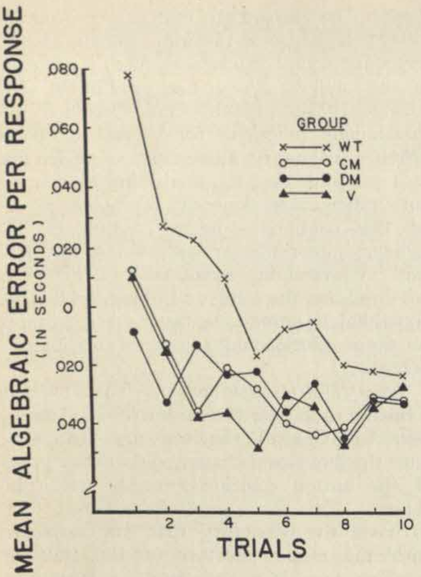


FIG. 1. Performance on the continuous pursuit tracking apparatus used as the transfer task in Exp. I.

Figure 1 is a plot of mean algebraic error for the 10 whole-task trials. The pretraining activities had a substantial positive transfer effect on whole-task performance because the three experimental groups all have less mean error throughout than the Group WT. Using total algebraic error for the 60 responses as *S*'s score on Trial 1, simple analysis of variance for all groups gave an *F* ratio of 4.45 ($df = 3/43$, $P < .01$). Tukey's gap test (Snedecor, 1956, p. 251) showed that the experimental groups did not differ significantly among themselves (.05).

The effect of the pretraining was to reduce the lag error which is so prominent for Group WT on Trial 1. Of course the mean algebraic error in the vicinity of zero for the three experimental groups could come from an appropriate distribution of relatively large plus and minus values without any beneficial anticipation whatsoever. To check this, we defined beneficial anticipation as a response to directional change occurring in the range of ±76 msec. This range was chosen because the smallest unit of measurement on our recording paper was 1 mm. which, at the chart speed we used, was 76 msec. Responses whose values were −76, 0, or +76 msec. were categorized as beneficially anticipatory because they are clearly less than any known human RT value. Our next level of response categorization would have been ±2 mm. or ±152 msec. but, since RT under some conditions can be less than 152 msec. we would not have been so confident of an index of beneficial anticipation had this gross criterion been used. The percentages of beneficial anticipations, based on the total number of responses made by each group on Trial 1, were greater for pretraining groups. The percentages for Groups WT, V, DM, and CM were 38, 52, 55, and 55, respectively. For statistical testing, a beneficial anticipation score was defined for each *S* as the number of beneficial anticipations on Trial 1. These scores for the four groups were significantly different at the .05 level ($F = 3.41$, $df = 3/43$). The gap test found no difference among Groups V, DM, and CM. We conclude that the Trial 1 differences for algebraic error in Fig. 1 are due mainly to beneficial anticipatory responses acquired in pretraining, but that pretraining did not function to differentiate among the three experimental groups.

EXPERIMENT II

Discrete tracking differs from continuous tracking in that each event is in a stable state. Anticipation of event change is based solely on time estimation, in contrast to continuous tracking where time estimation and changes in the signal pattern both occur as cues for anticipatory responses. Experiments II and III were run to broaden the scope of our findings in Exp. I. The results of Exp. II were negative and will be discussed succinctly as background for the positive findings of Exp. III.

Method

The Discrete Tracking Apparatus (DTA) was used, and is described under Exp. III. The stimulus sequence was a repetition of three events. The number of pretraining and whole-task trials, as well as trial duration and intertrial rests were the same as Exp. I. A control group (WT) had just whole-task visual discrete tracking. Four experimental groups in the pretraining trials either made a verbal response to predict the onset of each of three visual signals (VV), a verbal response to each of three auditory signals (AV), a visual-motor response of pressing one of three buttons to coincide with the onset of each visual signal (VM), or the same button pressing motor responses to auditory signals (AM). The *S*s were told that their pretraining performance was being scored. Each group had 8 male *S*s.

Results

The left-hand side of Fig. 2 shows the results in terms of time on target for the whole-task trials. While time on target ordinarily is used as an overall measure of motor skill in a tracking task, in discrete tracking *S*'s proficiency is directly a function of timing an elementary positioning response to a stimulus change, with none of the elaborate response sequences required in continuous tracking. With a very simple repetitive stimulus sequence and with only rare wrong movements, as was the case here, off-target time is a matter of the accuracy of timing the response to each stimulus change, and so time on target becomes a direct index of anticipation. Perfect anticipation of each event would give 100% time on target. A simple analysis of variance test of time-on-target scores on Trial 1 lacked significance at the .05 level.

A beneficial anticipation score was

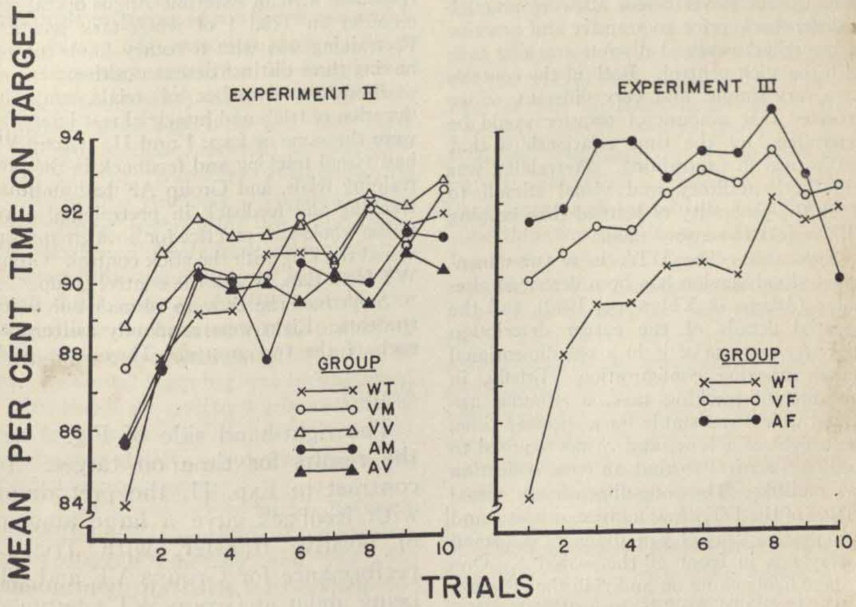


FIG. 2. Performance on the visual discrete tracking apparatus used as the transfer task in Exp. II and III.

determined for each *S* on Trial 1 of the whole-task, and a simple analysis of variance test of them for the five groups failed the .05 level. The chart units of the operations recorder used to record all stimulus and response events on Trial 1 were slightly different than in Exp. I and necessitated a criterion of ±67 rather than ±76 msec.

EXPERIMENT III

Method

We surmised that the negative findings of Exp. II were ascribable to poor knowledge of results. There was no feedback for any of the pretraining responses and it would be reasonable to expect that a mediating response as an explanation of perceptual anticipation would require knowledge of results for its learning. Experiment III was designed to see if anticipatory behavior is a function of feedback and, toward this end, some departure from the designs of Exp. I and II was used. Instead of having a neutral nontracking response for pretraining, pretraining in Exp. III was with a different version of the discrete tracking task so that feedback on the timing of responses could be had. The control was completely different in pretraining, but nevertheless allowing practice with feedback prior to transfer and practice in the criterion visual discrete tracking task with the stick control. Both of the controls were very simple, and very different, so we expected that amount of transfer would be determined by the time relationships that are learned in pretraining. Pretraining was with both auditory and visual stimuli to check the generality of learned time keeping with respect to sensory mode.

Apparatus.—The DTA in a two-dimensional visual version has been described elsewhere (Adams & Xhignesse, 1960), and the essential details of the earlier description apply for our use of it in a one-dimensional pursuit tracking configuration. Briefly, in this discrete tracking task, a stimulus appeared which was stable for a specified time, like a light or a tone, and *S* was required to position a control so that an error indication was nulled. The one-dimensional visual version of the DTA had a green, a white, and a red jeweled lamp of ½-in. diameter in a small display box in front of the seated *S*. One of these lights came on and *S* in the criterion task had to position a 2-in. control stick on the armrest of his chair so that a small feedback light beneath the cue light was on, signifying proper positioning of the control. If *S* was off target in one of the two wrong positions, the feedback light was under a cue light which was not illuminated (error cue) and he had to correct the error by moving his control to place the lighted feedback lamp under the cue light which was on. The cue light kept changing and the task of *S*s was to keep error to a minimum by having the feedback light in alignment with the cue light as much as possible.

In an auditory tracking version of this task three tones were presented sequentially over a headset, either 600, 800, or 1000 cycles. When *S* was on target he heard a pure tone, but when he was off target he heard a complex tone made up of a pure tone signifying the correct signal and, superimposed on it, one of the other two signals, rapidly interrupted, indicating the misalignment of the control.

The input series were regular repetitions of the three 1-sec. signals and were programed on punch tape and read by an automatic tape reader. Successive events were perfectly contiguous in time. This required a regular series of three discrete movements from *S*, back and forth, and beneficial anticipation was revealed in the closeness of the stimulus change and the response to it. Performance measurement was time on target, as well as complete recording of all stimulus events and responses with an Esterline-Angus operations recorder on Trial 1 of whole-task practice. Pretraining was with a rotary knob control having three distinct detent positions.

Procedure.—Number of trials, and the duration of trials and intertrial rest intervals, were the same as Exp. I and II. Group VF had visual tracking and feedback in the pretraining trials, and Group AF had auditory tracking and feedback in pretraining. Criterion whole-task practice for both groups was visual tracking with the stick control. Group WT from Exp. II was the control group.

Subjects.—The *S*s were 16 male university students. Eight were randomly assigned to each of the two groups. They were paid.

Results

The right-hand side of Fig. 2 has the results for time on target. In contrast to Exp. II, the pretraining with feedback gave a large amount of positive transfer, with Trial 1 performance for Groups VF and AF being about at Group WT's terminal performance level. A simple analysis of variance for time-on-target scores on Trial 1 gave a significance level beyond .01 ($F = 13.04$, $df = 2/21$).

Using the same criterion for beneficial anticipation as Exp. II, the percentages of these precisely timed responses for Groups WT, VF, and AF on Trial 1 were 39, 67, and 73, respectively. A simple analysis of variance for beneficial anticipation scores gave an F ratio significant beyond the .01 level ($F = 11.09$, $df = 2/21$). The large positive transfer on Trial 1 was importantly a function of the relatively large number of beneficial anticipations. Tukey's gap test showed no significant difference between the two experimental groups.

DISCUSSION

Positive transfer of pretraining to the accurate and anticipatory timing of directional changes in continuous visual tracking, and to the anticipatory timing of event changes in discrete visual tracking, abundantly confirm the usefulness of the pretraining approach for the study of anticipation and encourages an interpretation in terms of mediated responses. Experiment III revealed that feedback is a necessary condition for acquisition of beneficial anticipation in our discrete tracking task, but it is not clear why deliberate feedback action was unnecessary for the transfer of anticipation in continuous tracking. Apparently the cues during pretraining for continuous tracking were distinctive enough for *S* to know the adequacy of his response timing, but not so in the case of the 1-sec. discrete events. And, providing feedback was present, the transfer to discrete visual tracking was independent of the visual or auditory stimulus mode used in pretraining. An acquired timekeeping capability is not wedded to the original sensory mode.

No differences were found between verbal and motor methods of pretraining for continuous tracking, and this provides us with no basis for deciding between a motor and a nonmotor locus of the mediators. While further research may uncover evidence on the locus of mediators, the absence of differences between verbal and motor methods suggests a new explanation in terms of a common type of movement-produced stimuli that accompanies all responses, whatever their locus. Proprioception is one candidate for the common factor. All of our pretraining responding had proprioceptive accompaniments, including verbal responses, and these proprioceptive traces from persisting fractional movements associated with pretraining responses may be the mediational basis of anticipatory timing in motor behavior. Time-varying proprioceptive traces may be the response-produced stimuli which function as the time perception mechanism to provide distinctive cues at time t for cuing a beneficial anticipatory overt response to appear at time $t + \Delta t$. Perhaps any pretraining response can provide transfer of anticipation in tracking through the mechanism of proprioceptive stimuli associated with it, and it does not particularly matter whether the response locus is motor-peripheral or verbal-central. An alternative common factor might be cognitive learning, and this interpretation could hold that all pretraining methods allowed *S* to observe the time patterning of events, and these operations of association in the perceptual field, along with confirmation of the sign-significate expectation (Tolman, 1958), would seem to be a sufficient condition for acquiring anticipatory behavior. Perhaps it would have been useful to have a pretraining condition for cognitive learning where *S* merely watched the events, but we rejected this at the outset because such operations do not prevent the verbally facile human *S* from covert verbal responding, thus producing functional equivalence to our explicit verbal pretraining. We thought it better experimental control if our verbal responses were deliberately overt and observable.

We do not deny the possibility of cognitive learning as the primary basis of pretrained anticipation in our studies, but there is the scientific problem of unequivocally specifying operations which disallow the influence of verbal responses and their stimuli, proprioceptive or otherwise. Proprioception, on the other hand, is not quite so difficult to pursue experimentally because it can be manipulated through several defining variables (e.g., Bahrick, 1957). While evidence on uncontaminated cognition will be hard to obtain directly, a body of indirect evidence in its favor might be assembled if research revealed that the time-varying qualities of proprioception do not influence the accuracy of perceptual anticipation. If there is anything to this role for proprioception, anticipation in tracking should emerge as (*a*) some function of the physical variables of the control system associated with proprioceptive feedback, and (*b*) time between responses. Goldstone, Boardman, and Lhamon (1958) offer encouragement for the proprioceptive view with their preliminary evidence showing that children make extensive use of kinesthetic cues in originally learning to perceive time intervals.

SUMMARY

Three experiments were performed to test the hypothesis that anticipation of directional signal change in one-dimensional continuous tracking, and time of event change in discrete tracking, is some function of mediated responses and their response-produced cues. A transfer of training design was used where practice on the criterion tracking task was preceded by a pretraining session where neutral responses were made to time regularities in the task stimuli. The design also attempted to make a first delineation of the motor vs. nonmotor locus of mediating response determinants of anticipation, if evidence was found for mediation at all.

Significant amounts of transfer of training were found and mediation as a mechanism for anticipation in tracking was found tenable. No differences were found for the motor vs. nonmotor locus of the mediators. Time-varying proprioceptive traces and cognitive learning were discussed as explanations.

REFERENCES

Adams, J. A. A source of decrement in psychomotor performance. *J. exp. Psychol.*, 1955, **49**, 390–394.

Adams, J. A. Human tracking behavior. *Psychol. Bull.*, 1961, **58**, 55–79.

Adams, J. A., & Xhignesse, L. V. Some determinants of two-dimensional visual tracking behavior. *J. exp. Psychol.*, 1960, **60**, 391–403.

Arnoult, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 339–350.

Bahrick, H. P. An analysis of stimulus variables influencing the proprioceptive control of movements. *Psychol. Rev.*, 1957, **64**, 324–328.

Bartlett, F. C. Anticipation in human performance. In, *Essays in psychology: Dedicated to David Katz.* Uppsala: Almquist & Wiksells, 1951. Pp. 1–17.

Gibbs, C. B. The continuous regulation of skilled response by kinaesthetic feedback. *Brit. J. Psychol.*, 1954, **45**, 24–39. (a)

Gibbs, C. B. Movement and force in sensorimotor skill. In W. F. Floyd & A. T. Welford (Eds.), *Human factors in equipment design.* London: Lewis, 1954. Pp. 103–117. (b)

Goldstone, S., Boardman, W. K., & Lhamon, W. T. Kinesthetic cues in the development of time concepts. *J. genet. Psychol.*, 1958, **93**, 185–190.

Goss, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62**, 20–31.

Helson, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, **42**, 473–497.

Poulton, E. C. The basis of perceptual anticipation in tracking. *Brit. J. Psychol.*, 1952, **43**, 295–302. (a)

Poulton, E. C. Perceptual anticipation in tracking with two-pointer and one-pointer displays. *Brit. J. Psychol.*, 1952, **43**, 222–229. (b)

Poulton, E. C. On prediction in skilled movements. *Psychol. Bull.*, 1957, **54**, 467–478.

Snedecor, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State Coll. Press, 1956.

Tolman, E. C. *Behavior and psychological man.* Berkeley: Univer. of California Press, 1958.

Vanderplas, J. M. Transfer of training and its relation to perceptual learning and recognition. *Psychol. Rev.*, 1958, **65**, 375–385.

Vince, M. A. The part played by intellectual processes in a sensori-motor performance. *Quart. J. exp. Psychol.*, 1953, **5**, 75–86.

Vince, M. A. The relation between hand movements and intellectual activity in a skilled task. *Quart. J. exp. Psychol.*, 1955, **7**, 82–90.

(Received October 28, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 91–97

SHAPE-CONSTANCY: DEPENDENCE UPON STIMULUS FAMILIARITY[1]

C. ROBERT BORRESEN AND WILLIAM H. LICHTE

University of Missouri

The purpose of the present study was to specify, for one set of conditions, the role of learning in shape-constancy. Undertaking such an experiment does not mean that the present writers minimize the importance of stimulus variables in shape-constancy, nor does it imply that learning must be a major factor. The writers are impressed by the theory of Gibson (1950, 1959) and feel that this theory itself makes the present study desirable.

The intent of the study was not to attack the whole question of how learning influences perceptual constancy; it was limited to showing how one specific learning factor operated in one situation. Learning in this instance meant familiarity and the main feature of the study was that familiarity was manipulated as an experimental variable. The primary hypothesis tested was that the degree of shape-constancy shown in the responses to a given stimulus is a function of the amount of experience *O* has had with that stimulus. Experience (or familiarity) was crudely defined as the number of instances of being-stimulated-by-and-making-a-judgmental-response-to the object. The judgmental response arbitrarily selected was that of indicating the orientation of the stimulus. The perception of orientation (tilt or slant) has always seemed, from a logical standpoint, to be crucial in producing constancy (Beck & Gibson, 1955; Stavrianos, 1945).

On the assumption that the basic hypothesis would be supported, the experiment was also designed to evaluate the role of one condition—the *variety* of experiences with a form—in the familiarizing process. Shape-constancy is essentially the phenomenon of perceiving the same "real" shape as the retinal image pattern changes because of the different orientations of the stimulus object, therefore the *number* of orientations in an experience period could be important. The second hypothesis was that the degree of shape-constancy would be a function of this variable.

The general plan of the experiment was to use initially unfamiliar ("nonsense") figures and to have observers view-and-respond-to these different numbers of times in an experience or familiarity period. Shape-constancy for each stimulus was then to be measured in a test period. To test the second hypothesis, different groups of *S*s were to see the stimuli at different numbers of orientations in the experience period.

METHOD

Apparatus.—A sketch of the apparatus, as seen by *O*, appears in Fig. 1. The apparatus was designed to present the standard stimulus (SS) at various slants directly in front of *O* (B in Fig. 1), at a distance of 1.22 m. from his forehead; the variable stimulus (VS) used for constancy judgments was 30° to the right of the SS position (C in Fig. 1), and the circular stimulus used for orientation judgments during training was 30° to the

---

[1] This research was supported in part by Contract AF33(600)31315, held by International Business Machines Corporation, University of Missouri subcontract.

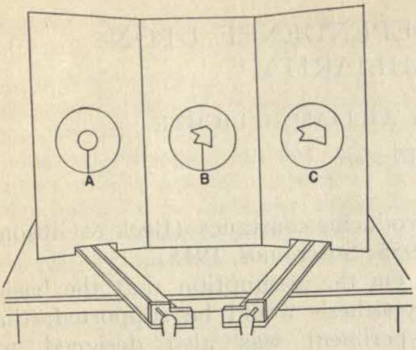


FIG. 1. Sketch of apparatus, as seen by the observer.

left of the SS position (A in Fig. 1). The *O*'s head was the center of a circle with points A, B, and C on its circumference. The *O*'s head was not restrained, although he was instructed to maintain a specified position. The *O* always used binocular vision.

For convenience, the word "stimulus" will be used in this paper to refer to the "distal" stimulus. Further, as will be seen below, the SS in this study was always a physical object (a piece of paper with a backing) while the VS was always a patch of light seen on a screen.

The *SS* shape in the middle position could be seen through a circular opening, 18 cm. in diameter, which had a removable shield between the opening and the stimulus. The shield was put in place between trials so that *O* could not see *E* changing the stimuli. A 7-w. bulb illuminated the opening so that *O* could clearly see the stimulus in the semidark room. The illumination on the SS was about 5 ft-c. Each SS was made of a piece of heavy photographic film faced with white paper.

On *O*'s right was the VS mechanism. The VS was produced by passing a beam of light through a piece of black photographic film with an appropriate transparent shape in the center. The beam of light then fell on a paper screen (from the rear). Light was provided by a Project-O-Chart (American Optical Company) with a 50-w. bulb, 4.5 m. from the screen. This method only approximated parallel light, but the resulting figures were not appreciably distorted or blurred. The VS mechanism was boxed in to prevent any stray light from striking the screen. The light beam could be blocked off by a shield between trials. The brightness of the patch of light constituting the VS was approximately equal to that of the illuminated white paper of the SS.

The film rectangles for producing the variable stimuli were rigidly mounted in frames which could be slipped into a holder on top of a rod in back of the paper screen. The *O* could rotate the rod and thus the film by turning a crank in front of him. By means of this rotation of the film the VS could be changed from one duplicating the shape on the film to a very narrow projection of this shape. A pointer attached to the rod supporting the film rotated over a fixed protractor so that the position of the film could be read by *E*.

On the left of *O* was a stimulus for orientation judgments obtained during the familiarity period. This consisted of a circular shape, of about the same area as the average of the standard stimuli, which could be rotated by *O* on a vertical axis through its center. A 7-w. light source illuminated this shape. Again, the illumination on the white plate was about 5 ft-c.

Stimuli.—Since the purpose of the experiment was to manipulate familiarity (number of past stimulus-response sequences), it was not possible to use the geometrical shapes usually employed in shape-constancy studies because of the previous experience of *O*s with such shapes. New shapes were therefore constructed for this experiment. They were

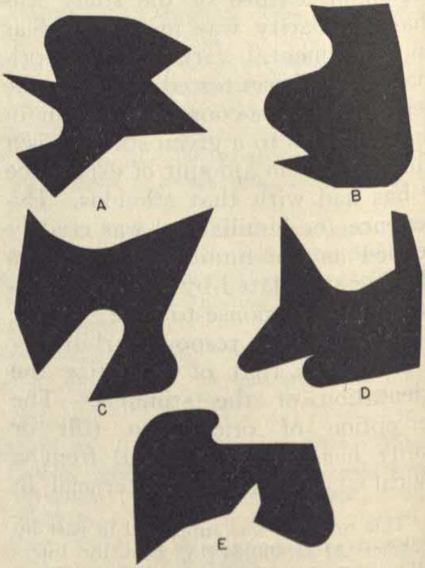


FIG. 2. The standard stimuli.

created by using a very slight modification of one of the methods (No. 4) suggested by Attneave and Arnoult (1956). The shapes employed are presented in Fig. 2.

As described previously, the VS was produced by passing a beam of light through a film. On this film the shape portion was transparent, while the surround was not. This allowed for a white projection on a black (or very dark) field. In all cases the shapes on the film were the SS shapes expanded horizontally, but not vertically. The difference was such that when the film was placed at an angle of approximately 60° from the path of the light beam the VS (on the screen) matched perfectly the shape of the corresponding SS (in the frontal plane). This relationship (film to image) was used for two reasons: (*a*) it allowed for the measurement of overconstancy, if it occurred (the VS could be wider than the physical object which was the SS); and (*b*) it avoided a suggestive perceptual landmark on the VS continuum. If the shape on the film had been the same as the SS, the VS on the screen would, as *O* turned the crank, expand to the point of equality with the objective shape of the *SS* and then begin to contract. This point of change would represent perfect shape-constancy.

Subjects.—The *O*s were all from undergraduate psychology classes at the University of Missouri. There were no restrictions made with regard to wearing glasses, age, or sex. There were 40 *O*s in all, 15 in each of the two experimental groups and 10 in a control group.

Procedure.—There were two parts to the experiment, a familiarity period and a constancy period. During the first, *O*s made judgments about the orientation of the SS; during the second they made shape-constancy judgments.

There were two experimental variables in the familiarity part of the experiment: (*a*) frequency of presentation, and (*b*) the orientation of the shapes.

Each *O* had 40 familiarity trials in which the five stimuli were presented at different frequencies. For any given *O* the five stimuli were presented at frequencies (one stimulus at each frequency) of 0, 4, 8, 12, and 16. There were five separate combinations of frequency-of-presentation and shapes; in other words, all shapes were rotated through all frequencies of presentation in a latin square design.

The second variable in the training period was the orientation of the *SS*. (The orientation or position of the stimulus was measured with respect to the line of sight from *O* to the stimulus, 0°. When the stimulus was in the frontal-parallel plane the edge to the left of *O* was at 90°.) In terms of the orientation variable there were two groups of *O*s. In Group I *O*s had one-half the presentations of each shape at 110° and one-half at 140°. Group II had one-fourth of the presentations of each stimulus at 110°, one-fourth at 120°, one-fourth at 130°, and one-fourth at 140°. This design allowed the absolute frequency of any particular stimulus to be the same under all conditions. For example, if Stimulus B were to be presented eight times, it would appear in two orientations (110° and 140°) four times each in Group I, but in four orientations (110°, 120°, 130°, and 140°) only two times each in Group II.

A control group of 10 *O*s (Group III) did not have any familiarity trials.

On each familiarity trial *O* was to indicate the orientation of the SS by rotating the circle on his left until he thought that it matched perfectly the orientation of the SS. This judgment was independent of shape, as the standard stimuli were the nonsense shapes whereas the orientation shape was a circle. The order of presentation of the standard stimuli was randomly determined for each condition and that order, and its inversion, were used in a counterbalanced manner.

The essentials of the instructions given *O* at this point were that the experiment involved judgments about abstract shapes, that various shapes would be placed in the middle window at some angle of slant, and that the task was to adjust the slant of the circle in the left window until it duplicated that of the object in the center window. The *O* was to face towards a spot halfway between the two windows and keep the head in that position.

After the familiarity period *O* had a few minutes rest, and then went on to the shape-constancy part of the experiment. Each *O* (including the control *O*s) had 30 constancy trials, 6 for each of the five stimuli. Each SS was placed at an orientation of 135° for three trials and at 155° for three trials. The use of different orientations was primarily to prevent *O*s from becoming too fixed in their responses by recognizing that the SS always appeared in the same orientation. These particular values were chosen on the assumption that they would normally (with familiar figures) give Brunswik ratios in the upper-middle range (Lichte, 1952).

It was *O*'s task during the constancy trials to view the rotated SS and adjust the VS film plate until he thought that the VS image on the screen was the same shape as the SS when at 90°, i.e., the objective shape of the SS. The *O*s rotated the VS film, thus changing the VS image, by turning a crank. For one-half of the trials of each shape the VS was set to give an image narrower than the frontal plane projection of the SS, and one-half the trials, wider. After *O* had made an adjustment to his satisfaction his score was read from a pointer attached to the VS rod. During the constancy trials, two different random orders of presentation of the SS were used in a counterbalanced manner.

The substance of the instructions given *O* for the second part of the experiment was that he would see various shapes (objects) in the middle window, and in the right window he would see patches of light on the translucent screen. The patch of light could be changed in shape by turning a crank and *O*'s task was to adjust its shape to duplicate that of the plate in the middle window, when the latter was considered as an object independent of its slant; *O* was to face midway between the two windows.

RESULTS AND DISCUSSION

Transformation of scores.—As the film was manipulated the shape of the VS on the screen changed continuously in width, but remained the same in height. Each shape on the continuum was associated with one image width. The measurement of image width was therefore considered a satisfactory index of the physical variable of shape. The original data for *O*s' responses were in terms of the angle of the film plate with reference to a line perpendicular to the screen. By means of the appropriate trigonometric functions these angles were transformed into width measures for the VS on the screen.

There were three trials for each SS presentation and therefore three width scores. The means of these widths were used to compute a BR (Brunswik ratio). There were 10 BRs for each *O*. These data were treated by an analysis of variance. The particular test used was a combination of a factorial and latin square design.

TABLE 1

ANALYSIS OF VARIANCE OF BRUNSWIK RATIOS

Source	*df*	*MS*	*F*
Between *S*s	29	.23	
Orientation	1	.01	
$S \times F_{(b)}$	4	.42	1.91
$S \times F \times O_{(b)}$	4	.15	
$Error_{(b)}$	20	.22	
Within *S*s	120	.11	
Shapes	4	1.02	20.40**
Frequency	4	.43	8.60**
$S \times F_{(w)}$	12	.11	2.21*
$S \times O$	4	.02	
$F \times O$	4	.11	2.20*
$S \times F \times O_{(w)}$	12	.18	3.60**
$Error_{(w)}$	80	.05	

* $P < .05$.
** $P < .01$.

The complete analysis of variance is presented in Table 1. Frequency of presentation as a main effect (*F*) was significant at better than the .01 level. The mean BRs for frequencies of 0, 4, 8, 12, and 16 were .46, .66, .61, .77, and .74, respectively. These results in general support the main hypothesis that shape-constancy is an increasing function of shape familiarity where familiarity is defined as the number of times that a shape was seen and responded to by *O*. This function, separated out for Groups I and II, is presented in Fig. 3.

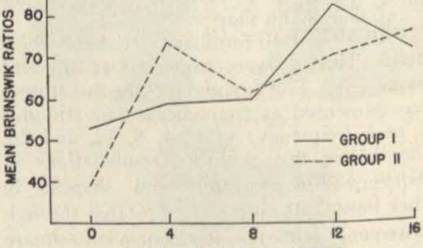


FIG. 3. Brunswik ratios for each frequency, by group.

The writers ascribe considerable theoretical importance to this basic finding. The function obtained is considered important, but so also is the obtained BR of .46 for the zero frequency-of-presentation condition. It seems reasonable to conclude from this that the stimulus factors discussed by Gibson (1950, 1959) are operating to produce this much shape-constancy.

A comparison of the mean BR of .43 for the control group of 10 *O*s (with no familiarity trials) with the zero frequency-of-presentation condition for the experimental groups indicated that there was no significant difference between the two conditions. This result is interesting in that it may be interpreted as evidence against any generalization or positive transfer in the experimental groups from the four stimuli in the familiarity period to the fifth stimulus, seen first in the second period. If there had been positive transfer one would expect the zero frequency-of-presentation condition of the experimental groups to have a somewhat higher BR than the control group.

An analysis was run comparing the 4, 8, 12, and 16 levels of frequency of presentation. The results of this test indicated no significant difference between these levels of presentation. The *t* test was applied to all possible pairs of conditions and only the 0–12 and 0–16 differences were significant at the .01 level. This may mean that the zero frequency-of-presentation factor is contributing the major portion of the variance which resulted in frequency of presentation being a significant variable. This function deserves further study, however, for in a nearly-identical previous study (Borresen & Lichte, 1959) with 30 *O*s a significant function was found when the frequencies of presentation were 2, 4, 8, 12, and 16.

The number of orientations (0) of the SS during the familiarity session was not significant; the mean BR for Group I was .65 and for Group II, .64. A glance at Fig. 3 clearly indicates the extensive overlap between these two groups. Thus, the secondary hypothesis, that shape-constancy increases with the number of different orientations, was not supported. However, the conditions of the experiment gave a very limited approach to the subject of orientation and the relations here seem worthy of further study.

Shape as a main effect (S) was significant at better than the .01 level. For Stimuli A, B, C, D, and E the BRs were .63, .50, .51, .65, and .95, respectively. A subsidiary analysis comparing shapes A, B, C, and D indicated no significant difference between these shapes. It is of additional interest to note that in the control group also, Stimulus E had a very high BR. In this group the mean BRs for Stimuli A, B, C, D, and E were .28, .33, .29, .42, and .80, respectively.

This points up a rather interesting situation. The shapes used in this study were all constructed by the same set of rules; they are, in a sense, all from the same physical family. Yet the data indicate that psychologically they are not the same. Other studies, with dependent variables such as shape recognition or discrimination (Arnoult, 1954; Fitts, Meyer, Rappaport, Anderson, & Leonard, 1956), have pointed toward the same conclusion: that shapes of a given physical family are not necessarily of the same psychological family.

The importance of the shape characteristic of the stimuli and the

TABLE 2

MEAN BRUNSWIK RATIO FOR EACH SHAPE AT EACH FREQUENCY

Shapes	Frequencies				
	0	4	8	12	16
A	.38	.56	.60	1.02	.60
B	.41	.47	.43	.36	.82
C	.55	.44	.43	.52	.62
D	.19	.86	.58	.79	.85
E	.77	1.00	1.01	1.17	.81

complexity of the whole subject of the relation of constancy to familiarity is indicated by the fact that the Shape × Frequency interaction is significant at less than the .05 level. The means relating to this interaction are shown in Table 2. A great deal of chance variability is undoubtedly represented here, perhaps particularly in the case of Stimulus D, but the table suggests the importance of further study of this interaction.

The perceptual constancies are undoubtedly most complete in realistic or everyday situations. In this study the stimulus situation was artificial and rather impoverished and thus might be expected to produce incomplete constancy. In addition, the nature of the instructions to *O* was not such as to produce the greatest degree of constancy. The latter conclusion comes from a subsequent study on the effect of instructions (Lichte & Borresen, 1960). As a result, at least in part, of these conditions incomplete constancy was found and this may have increased the probability of finding, or the size of, the hypothesized relationship between familiarity and degree of constancy.

In other studies with somewhat different conditions (Sheehan, 1938; Stavrianos, 1945) high degrees of constancy or overconstancy have been obtained. These prove little with regard to the point at issue, however, because the stimuli were rectangles, ellipses, and diamonds, and were therefore highly familiar at the outset of the experiment.

SUMMARY

The present experiment tested the hypothesis that shape-constancy is a positive function of shape familiarity, where familiarity is defined as the number of times *O* is allowed to view-and-respond-to a shape during a familiarity period. Some effects of the orientation of the standard stimulus during a familiarity period were also tested.

The stimulus materials were five moderately complex shapes systematically constructed from randomly determined coordinates. During a familiarization period the five stimulus shapes were presented 0, 4, 8, 12, and 16 times, one stimulus at each frequency, and *O*s were asked to indicate the orientation of the SS by matching a circular shape to that orientation. There were two familiarization conditions, each having a different number of orientations of the SS (two for Group I, $N = 15$, and four for Group II, $N = 15$). A control group ($N = 10$) received no familiarity trials.

In the second part of the experiment constancy was measured, using the same five forms. The *O*s were asked to set the VS to match the real shape of the SS. Two orientations were used.

The results supported the main hypothesis that shape-constancy is a positive function of shape familiarity. Shape-constancy for the zero frequency-of-presentation condition for the experimental groups was not significantly different from that of the control group. Shape as a main effect was significant, with most of the variance coming from one shape, and the interaction of shape and frequency was significant. Orientation during the familiarity period was not significant. The number of orientation positions, within a given range, seems to be unimportant.

REFERENCES

ARNOULT, M. D. Shape discrimination as a function of the angular orientation of the stimuli. *J. exp. Psychol.*, 1954, **47**, 323–328.

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.

BECK, J., & GIBSON, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50**, 125–133.

BORRESEN, C. R., & LICHTE, W. H. Shape constancy as a function of form familiarity.

Paper read at Midwestern Psychological Association, Chicago, May 1959.

Fitts, P. M., Meyer, W., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1956, **51**, 1–11.

Gibson, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950,

Gibson, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 456–501.

Lichte, W. H. Shape constancy: Dependence upon angle of rotation; individual differences. *J. exp. Psychol.*, 1952, **43**, 49–57.

Lichte, W. H., & Borresen, C. R. The influence of instructions on degree of shape constancy. Paper read at Midwestern Psychological Association, St. Louis, May 1960.

Sheehan, M. R. A study of individual consistency in phenomenal constancy. *Arch. Psychol.*, Geneve, 1938, **31**, No. 222.

Stavrianos, B. K. Relation of shape perception to explicit judgments of inclination. *Arch. Psychol.*, Geneve, 1945, **42**, 1–94.

(Received October 31, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 98–104

LATENT EXTINCTION AS A FUNCTION OF NUMBER AND DURATION OF PRE-EXTINCTION EXPOSURES[1]

JAMES A. DYAL[2]

Southern Methodist University

The latent extinction effect is a decreased resistance to extinction resulting from direct placement of *S*s into nonreinforced contact with previously positive stimuli prior to regular extinction trials. The weight of the evidence provides convincing support for the reliability of the effect since 10 out of 12 experiments have yielded positive results. In addition to demonstrating the effect, a few of these studies have attempted to determine the influence of other variables on latent extinction. Moltz and Maddi (1956) investigated the drive variable at three levels of hunger (0, 22, and 44 hr.). Their data suggested a negatively accelerated function relating drive level and number of correct choices during extinction. Pliskoff (1955) obtained similar results using 0, 8, and 22 hr. of hunger. He also investigated number of latent extinction exposures (1, 4, and 20) and found that this variable influenced choice behavior only if the placements were given in the same goal box as used during training. A parametric study by Thomas (1958) explored the effects of duration of a single goalbox placement (1, 5, or 15 min.) and interval between goalbox placements and extinction (15 sec., 5 min., or 15 min.). The purpose of the present experiment was to investigate the effects of duration and number of goalbox placements on the latent extinction effect.

METHOD

Subjects and apparatus.—The *S*s were 96 naive Sprague-Dawley male rats between 80 and 100 days old at the start of training. The apparatus was a Y maze, 5¾ in. deep and covered with ¼-in. hardware cloth. Upon leaving the 11 × 4 in. neutral gray startbox, *S* encountered a choice area shaped as a truncated triangle and having the following dimensions: base, 24 in., top, 10½ in., and sides, 13¾ in. The runway arms were both 5½ × 4½ in. A left turn led to a flat white runway having 36 nickel-plated, round-head thumbtacks set haphazardly in the floor. The circular goalbox was also painted flat white and was 10½ in. in diameter. Its sides were constructed of stainless steel and the wooden floor was covered with a removable piece of ¼-in. hardware cloth. A right turn led through an unpainted runway to an unpainted triangular goalbox having 13-in. sides. The *S*s thus had available brightness, tactual, and extramaze cues to facilitate discrimination of the arms and goalboxes. The floor of each goalbox contained a ½-in. deep food well which prevented *S* from seeing the food until he entered the goalbox. The neutral box into which the control *S*s were placed was 10 in. square and painted flat black. The apparatus was instrumented with a photoelectric system which permitted the recording of starting latencies, choice times, running times, and goalbox times on .001-min. Standard Electric timers. The goalbox times (GT) indicate the time required to traverse the last 2½ in. of the runway and the first 8½ in. of the goalbox.

Pretraining.—The *S*s were adapted to a 23-hr. food deprivation schedule for 7 days prior to the start of training. On Pretraining Day 5, *S*s were permitted to explore the apparatus in groups of 6 for 10 min. On Day

---

[1] The author wishes to thank Richard Carpenter and Bruce Patterson who ran the animals for this experiment. This research was supported by Public Health Service Research Grant M-3331 from the National Institute of Mental Health.

[2] Now at Texas Christian University.

6, each *S* was given 60 sec. contact eating following direct placement into each goalbox. The *S* was then given five free-choice trials with no food to determine turning preferences. On Day 7, each *S* received three massed forced choices to food on the nonpreferred side. Eating was allowed for 60 contact sec. on the forced trials.

Training.—The *S*s were trained in groups of 6 and were given five free-choice trials per day. The food (one brick of Purina laboratory chow) was always on the nonpreferred side. Following a correct choice, *S* was permitted to eat in the goalbox during the period when *E* was recording the time data. This period averaged 10–15 sec. The intertrial interval was never less than 5 min., which time *S*s spent in wire cages with water continuously available.

A multiple criterion of learning was established: (*a*) each *S* must have had a minimum of 25 training trials; (*b*) 9 of the last 10 trials must have been correct; and (*c*) the error if any, must not have been on the first trial of the last day of training.

Prior to reaching criterion, each *S* was randomly assigned to one of 12 groups. Nine experimental *S*s in each replication were assigned to one of the combinations of duration and number of placements represented in Fig. 1. Three *S*s were assigned to control groups having the duration and number values of the diagonal cells. The experiment was run in eight replications by 2 different *E*s. For convenience of exposition, the specific combination of duration, number, and experimental vs. control will be designated by group numbers, e.g., Group 6 refers to *S*s receiving 10 placements of 60 sec. duration in the goalbox, whereas 5C refers to 5 placements of 60 sec. duration in the neutral box. The effects of the independent variables of number and duration are present in Cells 1 through 9 and this will be referred to as the 3 × 3 design. The presence of the latent extinction effect is represented in the experimental and control portions of Cells 1, 5, and 9; the 2 × 3 design.

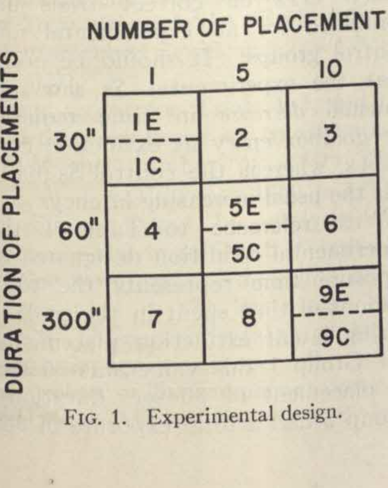


FIG. 1. Experimental design.

Latent extinction and regular extinction.—Twenty-two to 24 hr. after reaching the criterion each experimental *S* was given the appropriate number and duration of placements into the formerly baited goalbox. The interplacement interval of 60 sec. was spent in the retraining cages and in transit. Sixty to 90 sec. following the last goalbox placement, each *S* was given the first of 10 regular extinction test trials. The intertrial interval ranged from 5 min. early in extinction to 10 min. late in extinction. The procedure was the same for the control *S*s except they were placed into a neutral box rather than the goalbox.

RESULTS

In order to interpret the parametric data from the experimental *S*s in the 3 × 3 design as being relevant to latent extinction, it is first necessary to demonstrate that latent extinction actually occurred. The relevant data are provided by comparison of the experimental and control groups in the 2 × 3 design.

2 × 3 *Design*

Goalbox times.—On the basis of a pilot study it was determined that the dependent variable which was likely to be the most sensitive to the effects of goalbox placements was the goalbox time (GT) on the first correct trial.[3] Because of apparent heterogeneity of variance of the raw data, each GT score was transformed by log (1 + GT). Bartlett's test pro-

---

[3] The running times, choice times, and starting times indicated the same general trends as the goalbox times but the E-C differences were small and became smaller the further the segment was from the goal area. The data from these measures were not analyzed further.

TABLE 1

SIGNIFICANT RESULTS OF ANALYSIS OF VARIANCE FOR LOG (1 + GT) AND ERROR SCORES IN VARIOUS SEGMENTS OF EXTINCTION (2 × 3 DESIGN)

Extinction Phase	Source	*df*	*MS*	*F*
	Log (1 + GT) Scores			
1st correct trial	Exp. vs. Control	1	.3601	23.38****
Correct choices Trials 1–5	Exp. vs. Control	1	.1202	13.07****
	Exposure time	2	.2056	21.64****
	Exp. vs. Control × Exposure time	2	.0580	6.10****
Correct choices Trials 1–10	Exp. vs. Control	1	.0453	6.66***
	Exposure time	2	.0378	5.55***
	Error Scores			
Trials 1–5	Exp. vs. Control	1	.1480	7.18**
Trials 1–5	Exp. vs. Control × Exposure time	2	.0825	4.00*
Trials 1–10	Exp. vs. Control	1	.1660	4.61*
Trials 1–10	Exp. vs. Control × Exposure time	2	.1460	4.06**

* $P < .05$.
** $P < .025$.
*** $P < .005$.
**** $P < .001$.

vided no basis for the rejection of the hypothesis of homogeneity of variance with regard to the transformed data.

Results of analyses of variance applied to each of the extinction segments as well as to all 10 trials combined, are presented in the upper part of Table 1. Only those variables which contributed significant variance are included; also note that GTs are based on correct trials only. It may be seen that the experimental-control differences are highly significant during all extinction segments except Trials 6–10. Subgroup comparisons based on *t* tests in which the error term was derived from the mean square for error from the analysis of variance support the conclusion that the latent extinction effect is primarily related to 5E-5C and 9E-9C differences. In all extinction phases Groups 5E and 9E took significantly longer to enter the positive goalbox than the comparable control groups. The differences for Groups 1E and 1C, although in the appropriate direction, were not statistically reliable. The size of the latent extinction effect, although still significant when all 10 extinction trials are used as the unit of analysis, tends to diminish with continued extinction; there were no reliable differences between any of the experimental and control groups on Extinction Trials 6–10. This view is further supported by Fig. 2 in which GTs on correct trials are combined for all experimental and control groups. It should be noted that the experimental *S*s show an unusual *decrease* in time required for goalbox entry as extinction continues, whereas the control *S*s manifest the usual increasing latency.

With reference to Table 1 the experimental condition designated as exposure time represents the total amount of time spent in the goalbox during latent extinction placements. For Group 1 this value was 30 sec. (1 placement of 30 sec. duration); Group 5 had a total exposure of 300

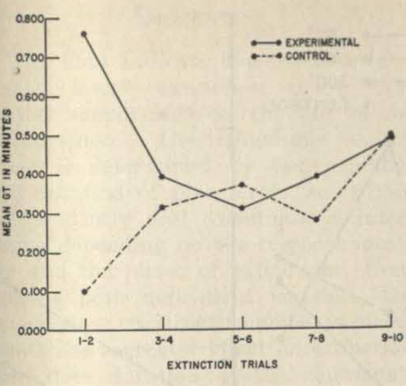


FIG. 2. Mean goal times for all experimental and all control groups combined, in blocks of two extinction trials.

sec. (5 × 60 sec.), and Group 9 had an exposure of 3000 sec. (10 × 300 sec.). Subgroup comparisons indicated that Group 1E entered the goalbox significantly faster than either Group 5E or Group 9E during the first 5 trials of extinction and over all 10 extinction trials. In no case was Group 5E significantly different from Group 9E.

Errors.—An error is defined as a choice of the previously unbaited goalbox. Evaluation of the number of errors made by the subgroups on Test Trial 1 by a test for uncorrelated proportions indicated that only the 1E-1C difference was reliable ($z = 2.00$; $P < .025$) although the 9E-9C difference approached significance ($z = 1.51$; $P < .07$).

Error scores for Trials 1–5, 6–10, and 1–10 were based on transformation of the number of errors by $\sqrt{x} + \sqrt{1 + x}$ (Edwards, 1960, p. 128) and were analyzed by analysis of variance. The significant F ratios resulting from each of these analyses are presented in the lower part of Table 1. The experimental vs. control variable is significant over all trials except 6–10. The significant interaction was due to the reversal in number of errors for Groups 1E and 1C.

Another index of performance based on error scores is the trial on which the first error was made. Analysis of variance applied to these scores transformed by $\sqrt{x} + \sqrt{1 + x}$ resulted in a highly significant difference between experimental and control conditions ($F = 22.73$, $df = 1/42$, $P < .001$).

3 × 3 *Design*

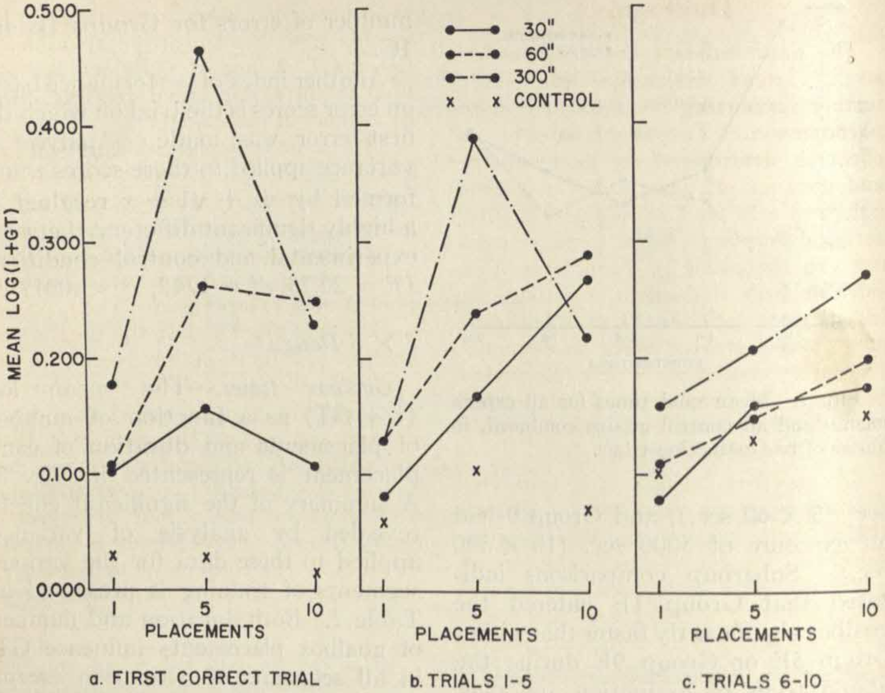
Goalbox times.—The mean log (1 + GT) as a function of number of placements and duration of each placement is represented in Fig. 3. A summary of the significant effects revealed by analysis of variance applied to these data for the various segments of training is presented in Table 2. Both duration and number of goalbox placements influence GT in all segments of extinction except Trials 6–10. Duration has its greatest effect on Trial 1 and does not influence Trials 1–5 except in interaction with number, whereas the main effect attributable to number of placements is present throughout the first 5 trials. It may be seen that the strength of the effect during Extinction Trials 1–5 is sufficiently great that when all 10 trials are combined both variables significantly influence GT.

TABLE 2

SIGNIFICANT RESULTS OF ANALYSIS OF VARIANCE OF LOG (1 + GT) FOR VARIOUS PHASES OF EXTINCTION (3 × 3 DESIGN)

Extinction Phase	Source	*df*	*MS*	*F*
1st correct trial	Duration (D)	2	.1684	5.36***
1st correct trial	Number (N)	2	.1704	5.34***
Correct choices Trials 1–5	Number (N)	2	.1553	25.88****
	N × D	4	.1544	25.73****
Correct choices Trials 1–10	Duration	2	.0910	12.13***
	Number	2	.0759	10.17***

*** $P < .005$.
**** $P < .001$.



a. FIRST CORRECT TRIAL b. TRIALS 1-5 c. TRIALS 6-10

FIG. 3. Mean log (1 + GT) during three extinction phases as a function of duration of placement.

Appropriate subgroup comparisons of the data represented in Fig. 3 and Table 3 reveals that Group 8 (5×300 sec.) was significantly ($P < .005$) slower in entering the goalbox on the first correct test trial than was Group 7 (1 × 300 sec.) or Group 9 (10 × 300 sec.). Group 8 was also significantly ($P < .001$) slower than Group 5 (5 × 60 sec.). During Trials 1–5 those groups which were given 30-sec. placements differed reliably as a function of the number of placements; Group 1 < Group 2 < Group 3 ($P < .025$, $P < .005$). Group 5 (5×60sec.) was significantly ($P < .005$) slower than Group 4 (1 × 60 sec.). For those groups having a duration parameter of 300 sec., Group 8 was reliably slower than Group 7 ($P < .001$) and Group 9 ($P < .001$). During Trials 1–10, Group 6 (10 × 60 sec.) was slower than Group 3 (10 × 30 sec.) and Group 8 was slower than Group 5 (both Ps $< .005$). Two significant differences were obtained for the number variable; Group 2 (5 × 30 sec.) was slower than Group 1 (1 × 30 sec.) and Group 8 was slower than Group 7 ($P < .05$, $P < .001$).

Errors.—Analyses of variance and t tests were applied to the transformed error scores for various phases of extinction. With the exception of a significant interaction with duration for the last half of extinction, it is only the number variable which significantly influenced choice behavior. Subgroup comparisons indicated that the effect was dependent primarily on the significantly greater number of errors made by Group 8 than by Groups 7 or 9.

DISCUSSION

The data indicate the presence of a strong latent extinction effect, thus further supporting the reliability of the phenomenon. The magnitude of the effect is determined by both number and duration of placements, sometimes independently and sometimes in interaction depending on the response measure and the phase of extinction. Considering both dependent variables, the independent variable of number of placements has a greater effect on extinction than does duration of the placement. However, the effects of both independent variables are transitory, being greatest during the first extinction trial.

The present results are somewhat opposed to those reported by Thomas (1958) with regard to the duration variable. He found that duration affected the error scores but not running times; whereas, in the present experiment GT was influenced by duration but error scores were not. If we assume that stimuli in closest association with the consummatory response will be most affected by nonreinforced placements then Thomas' choice measure is likely to have been more sensitive since his choice area was considerably closer to the goalbox than was the case in the present study (12 in. vs. 48½ in.). Similarly, however, the present speed measures (GT) should be more sensitive to the effects of goalbox placements than Thomas' running speed score. This view is supported by another experiment in the present series in which it was found that when measures of running speed were available for five segments of a straight runway it was *only* performance in the immediate goalbox area which was influenced by latent extinction procedures.

In the only other study dealing with the effects of number of placements, Pliskoff (1955) found that error scores were significantly influenced by number of placements throughout 20 extinction trials; whereas, in the present data the major effect of number was on the first 5 trials. He gave his *S*s 1, 4, or 20 exposures of 30-sec. duration. His groups are thus most comparable to Groups 1, 2, and 3 in which we also find an increasing trend relating errors to number of placements.

The most complete theoretical account of latent extinction per se has been provided by Moltz (1957) and is based on the diminution of the fractional anticipatory goal response resulting from nonreinforced elicitation of r_g during pre-extinction placements. The r_g framework would predict that the greater the number and duration of exposures in the empty goalbox the greater the reduction of r_g and the more *S*s would slow down in their approach to the goalbox. Also, the negatively accelerated relation between GT on the first correct trial and total exposure time would be expected by r_g theory. However, it is difficult to see how the concept of increasing r_g diminution could be invoked to account for the obtained *decrease* in GT exhibited by the experimental groups over the 10 extinction trials. On the contrary, r_g theory would predict a continued increase in running time as regular extinction proceeds.

The fact that the difference between experimental and control *S*s disappears by the end of 3–4 trials is quite consistent with a frustration induction theory of latent extinction. If we assume that nonreward in the presence of previously rewarded stimuli results in an emotional state of frustration, then this aversive drive should still be present 60 sec. after the last placement (Trial 1) and this would result in *S*s being reluctant to enter the goal area. However, such an emotional response would be expected to dissipate fairly quickly, perhaps in a matter of 15 min. (Thomas, 1958) after being returned to the home cage, and the resulting bow-shaped curve would be expected for the experimental groups. A more detailed examination of the performance changes from Trials 1–2 to Trials 3–4 indicates that the greatest shift occurred for those groups which had the greatest total exposure. This is also consistent with a frustration induction view in that the longer the exposure the greater the amount of frustration on Trial 1.

An implication of the present experiment for future studies of latent extinction is that the condition of five placements of 60-sec. duration (Group 5) is probably the condition of choice. It is economical of E time, and has the further advantage of being the approximate value on which the bulk of the previous literature is based.

SUMMARY

Ninety-six rats were trained to criterion in a Y maze and then given differential treatment with regard to number (1, 5, 10) and duration (30, 60, 300 sec.) of unrewarded goalbox placements. Control *S*s were run in the diagonal cells of the 3 × 3 design and differed from the experimental *S*s only in that they were placed into a neutral box.

A strong latent extinction effect was obtained for the dependent variables of speed of entering the goalbox (GT) and number of errors. The effect was strongest on the first correct trial in extinction and became progressively weaker as extinction continued. The magnitude of the effect was influenced by number of placements, duration of the individual placements, and the total exposure time, sometimes independently and sometimes in interaction depending on the response measure and the phase of extinction.

The results were compared with those obtained by previous investigators and two theories of latent extinction were examined. Both theories, frustration and r_g diminution, could accommodate the results pertaining to the effects of number and duration of placements. However r_g theory has difficulty explaining the transitory nature of the effect.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

MOLTZ, H. Latent extinction and the fractional anticipatory response mechanism. *Psychol. Rev.*, 1957, 64, 229–241.

MOLTZ, H., & MADDI, S. R. Reduction of secondary reward value as a function of drive strength during latent extinction. *J. exp. Psychol.*, 1956, 52, 71–76.

PLISKOFF, S. S. Response elimination as a function of generalization, motivation, and number of nonresponse extinction trials. Unpublished doctoral dissertation, New York University, 1955.

THOMAS, A. R. Some variables affecting latent extinction. *J. exp. Psychol.*, 1958, 56, 203–212.

(Received October 31, 1960)

ERRATUM

In the article "Role of Instructions in Two-Choice Verbal Conditioning with Contingent Partial Reinforcement" by J. Koehler, Jr., (*J. exp. Psychol.*, 1961, 62, 122–125), the line following the asymptotic expressions (p. 122) should read as follows: "It can be demonstrated algebraically that the first asymptotic expression is larger than the second one when $\pi_1 > \frac{\pi_2(1 - \pi_2)}{1 - \pi_1}$ and $\pi_1 < 1$."

Journal of
Experimental Psychology

Vol. 63, No. 2 February 1962

SOME VARIABLES INFLUENCING THE RATE OF GAIN OF INFORMATION[1]

ROBERT W. BRAINARD, THOMAS S. IRBY,[2] PAUL M. FITTS,[3] AND EARL A. ALLUISI[4]

Ohio State University

Choice time, or disjunctive reaction time (RT), has been found to be a linear increasing function of the average amount of information transmitted (H_t) per stimulus-response event in a variety of tasks (e.g., Gregg, 1954; Hick, 1952; Hyman, 1953). Hick found that the same function fitted the data when errorless performance was required and also when *S* speeded up his responses to the point where a substantial number of errors occurred. Hyman, demanding errorless performance, found that RT was a linear function of stimulus information, H_x, (and, therefore, of H_t) when three different procedures for varying average H_x were used: (*a*) the number of equally probable alternative stimuli was varied, (*b*) the relative frequency of occurrence of different stimuli was varied, and (*c*) the degree of sequential dependence of stimuli was varied. The mathematical equivalence of these three methods of varying stimulus uncertainty has been shown by Shannon and Weaver (1949); their psychological equivalence can be inferred from the results of Hyman's study.

Similar results have been obtained with the use of self-paced serial-reaction tasks. For example, in a serial card-sorting task Crossman (1953) found choice time to be linearly related to the average information processed per response, and in a serial switch-moving task Archer (1954) found that the average time per response increased as a linear function of the relevant information presented in the stimulus (over the range from 1 to 4 bits/stimulus), but was independent of the irrelevant information presented (over a range from 0 to 2 bits/stimulus).

---

[1] This report is a consolidation of data from two master's theses by Brainard and Irby; these were completed in 1956 and 1957 as part of a general research program for which P. M. Fitts and E. A. Alluisi were responsible, and which was supported in part by the United States Air Force under Contract No. AF 33(616)-43 and -3612, with the Ohio State University Research Foundation, monitored by the Aerospace Medical Laboratory, Wright Air Development Command. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

[2] Now at Dunlap & Associates, Incorporated, Stamford, Connecticut.

[3] Now at the University of Michigan.

[4] Now at Emory University.

Although a linear relation between RT and H_t of the form $\mathrm{RT}=a+bH_t$ has been consistently obtained, the slopes and intercepts of the functions, i.e., the numerical values of a and b, have varied widely from one task to another. Thus it is clear that parameters other than the informational characteristics of the task are significant in determining RT, specifically in modifying the slope and intercept of the function.

The present study was planned to investigate some of the parameters that may be responsible for the differences in slope and intercept obtained in earlier studies. Specifically, the objective was to determine the effects of (*a*) two types of stimulus codes (numerals and light patterns), (*b*) two types of response codes (vocal and finger), and (*c*) self-paced serial vs. discrete-reaction tasks. The four combinations of stimulus and response codes were chosen to provide high and low levels of S-R compatibility. Three levels of stimulus uncertainty (1, 2, and 3 bits/stimulus) were obtained by varying the number of equally probable alternative stimuli. Only naive *S*s were used, each *S* serving under only one S-R coding condition and one level of stimulus uncertainty. Thus the experiment also provides further tests of the generality of the linear relation of RT to H_t.

METHOD

The study was conducted in two parts, which were identical except for the specific S-R pairings used. Pairings of (*a*) key pressing responses to spatially arranged lights and (*b*) vocal responses to numerals were used in the first part. It was predicted that these S-R ensembles should show high compatibility i.e., give relatively high values of H_t per second. Pairings of (*c*) key pressing responses to numerals and (*d*) vocal responses to spatially arranged lights were employed in the second part. It was predicted that the latter S-R ensembles would be relatively less compatible since neither physical correspondence nor overlearning operated in its favor. The data of the two parts of the experiment were collected at different times, but the same apparatus and procedures were used throughout and *S*s were drawn from the same population of university students.

Stimuli.—Two sets of stimuli were used: 0.25-in. high Arabic numerals and 0.5-in. diameter ruby lights. The numerals appeared as projections on the center of a 10-in. diameter circular display of opal glass (brightness of 2.5 ft-L). The lights (440 ft-L) were mounted on a vertically placed panel; they were arranged spatially in two semicircles to correspond with the placement of the finger tips. Each of the stimulus surfaces was located approximately 28 in. directly in front of the seated *S*.

The amount of information characterizing a stimulus event was varied by random selection of stimuli from among two, four, or eight equally likely alternatives (1, 2, or 3 bits/stimulus, respectively). One set of conditions consisted of either two, four, or eight Arabic numerals. At the lowest level of stimulus uncertainty the numerals used were 4 and 7; the numerals 3, 4, 7, and 8 were used at the intermediate level, and all the numerals from 2 through 9 were used at the highest level of uncertainty. The other set of conditions consisted of either two, four, or eight ruby-colored lights. For reference purposes the lights, from left to right, were assigned the numbers 2 through 9; the specific lights used in making up the three levels of stimulus uncertainty were those corresponding to the three sets of numeral stimuli. Any light not being used in a specific condition was replaced with a 0.5-in. diameter metal cap that was set flush with the plane of the panel.

Mode of information presentation.—Stimuli were presented both in a self-paced serial-reaction task and in a discrete-reaction task. In the self-paced task, the first stimulus appeared approximately 3 sec. after a "get ready" burst of broad-band noise was presented through the earphones worn by *S*. Each succeeding stimulus appeared 0.15 sec. after *S* had responded to the previous stimulus. The interval of 0.15 sec. represents the electromechanical switching time of the experimental apparatus; it was subtracted from the self-paced response times to yield the RTs. (No corrections were made for the rise time of the incandescent lights, lags of the voice key or time clocks, or differential brightness of the ruby lights and projected numerals.)

Discrete stimuli were presented at the rate of one stimulus every 10 sec. A red circle in the center of the display screen (or a green light in the center of the array of ruby lights) appeared together with a broad-band noise signal to provide a "get ready" indication during the 2 sec. immediately preceding each stimulus presentation. The stimulus remained on for 8 sec.; it was followed by the 2-sec. warning for the next stimulus, etc.

Responses.—For the key pressing task, a bank of 10 keys was arranged horizontally in two semicircles to correspond with a natural placement of the finger tips. These keys were numbered from left to right, 1, 2, 3, . . . 10. Wooden blocks were inserted under those keys for which no corresponding stimulus would appear. The fingers used in the key pressing responses for the three levels of stimulus complexity were the two index fingers (4 and 7) for the 1-bit/stimulus level, the middle and index fingers of both hands (3, 4, 7, and 8) for the 2-bit/stimulus level, and all but the little fingers for the 3-bits/stimulus level. When making these motor responses, *S*'s task was to depress the key that corresponded to the stimulus presented. Specific errors were recorded on an 8 × 8 electromechanical counter matrix.

In making vocal responses *S*'s task was to identify the stimulus by responding with its numerical label. The responses were monitored and errors recorded by *E*, who had in front of him a list of the series of stimuli being presented to *S*. The *S*'s responses were also tape recorded so that ambiguities in scoring could be resolved by replaying the record.

Procedure.—The study employed a 4×3×2 factorial design, consisting of the four S-R codes (light and numeral stimuli paired with vocal and key pressing responses), the three levels of stimulus uncertainty (1, 2, and 3 bits/stimulus), and the two task conditions (self-paced and discrete).

Different groups of 20 *S*s were assigned to each of the combinations of the four S-R pairings with the three levels of stimulus uncertainty (12 groups); however each *S* served under both self-paced and discrete-reaction conditions. Each *S* served for five trials lasting somewhat less than an hour in all. Trials 1, 3, and 5 (100 responses per trial) required self-paced serial reactions, whereas Trials 2 and 4 (50 responses per trial) required discrete reactions. The first self-paced trial was given for purposes of task familiarization and data obtained from it were omitted from analysis.

Each *S* was seated in a small experimental booth. Ambient lighting was furnished by a single miniature fluorescent light; brightness contrast was high and all stimuli were quite legible. Broad-band noise (approximately 60 db.) was presented continuously to *S* through their earphones during all self-paced trials in order to mask extraneous sounds. Earphones were also worn during the discrete-reaction trials, but masking noise was presented only during the 2-sec. interval just prior to the presentation of each stimulus.

Subjects.—The *S*s were 240 undergraduate students at the Ohio State University, 156 men and 84 women. In each part of the study *S*s were assigned to one of the six groups used in that part on the basis of order in reporting to *E*, with the restriction that each group contain 13 men and 7 women.

RESULTS

Three measures of performance were computed: (*a*) average reaction time (RT, in sec/response), (*b*) average information transmitted per response (H_t, in bits/response), and (*c*) average rate of information transmission per response (H_t/RT, in bits/sec).

Relation between RT and H_t.—Average RT and average H_t were computed from the 100 discrete reactions of each *S*; similar estimates were also computed from the last 200 self-paced serial reactions of each *S*. Scores for *S*s were then averaged to obtain group estimates. The relation between these average RTs and H_t's were examined for each of the eight functions obtained for the different combinations of the two modes of information presentation (discrete and self-paced reactions) and the four S-R pairings. The least-squares linear equations of best fit for the eight conditions are given in Table 1. For none of the conditions was the deviation from linearity statistically significant ($P > .05$ in each case).

An average Pearson *r* between RT and H_t was obtained for all *S*s who used a given S-R code under similar (serial or discrete) conditions. For one of the more compatible conditions (key pressing responses to spatially arranged lights) and for both of the less compatible conditions the correlations were high (the average correlation was +.84). Thus, for these conditions, approximately 70% of the variation in average RT is accounted for by the variations in H_t. However, the average correlation between RT and H_t was only +.34 for the group tested under conditions in which vocal responses were made to numerical stimuli; here the variations in H_t account for only about 10% of the variations in RT.

Effect of mode of information presentation.—The two modes of information presentation (discrete vs. self-paced serial reactions) produced relatively small differences in performance. The linear relations between RT and H_t were nearly identical within each of the four S-R pairings and the differences between the four pairs of discrete and self-paced conditions in terms of the correlations of RT and H_t were in no case statistically significant ($P > .05$ in each case). There were small differences between the discrete and self-paced serial reactions in both RT and H_t, however, as shown in Fig. 1.

TABLE 1

EMPIRICAL LINEAR EQUATIONS OF BEST FIT FOR AVERAGE RT, IN SEC/RESPONSE, AS A FUNCTION OF AVERAGE H_t, IN BITS/RESPONSE

S-R Pairing	Task	Equation
N-M	SP	$RT = .281 + .183(H_t)$
N-M	D	$RT = .317 + .180(H_t)$
N-V	SP	$RT = .485 + .011(H_t)$
N-V	D	$RT = .360 + .061(H_t)$
L-M	SP	$RT = .229 + .111(H_t)$
L-M	D	$RT = .182 + .136(H_t)$
L-V	SP	$RT = .267 + .205(H_t)$
L-V	D	$RT = .335 + .178(H_t)$

Note.—N = Numerals, L = Lights, M = Motor (key pressing), V = Vocal, SP = Self-paced, D = Discrete.

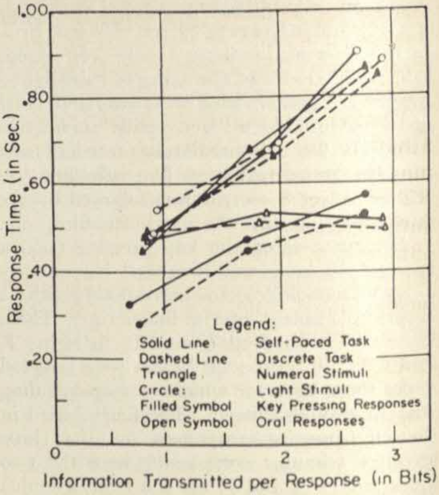


FIG. 1. Reaction time as a function of information transmitted for four S-R codes, and for self-paced vs. discrete tasks.

In the majority of cases the discrete reactions were made slightly slower and more accurately than the self-paced serial reactions; these changes tended to counteract each other, however, since average H_t/sec was almost identical for the two conditions.

S-R compatibility effects.—The most striking S-R compatibility effect is apparent from an examination of the slope constants for the linear equations of Table 1. In all cases the slopes of the functions relating RT and H_t for the less compatible tasks are greater than are those for the most compatible tasks. Thus, differences in RT attributable to the use of different S-R codes are present when only two alternatives are used, but increase markedly as number of alternatives (and H_t per response) increases.

DISCUSSION

The present results agree with those of previous investigations showing that RT is a linear function of H_t. They also agree with previous studies (cf. Alluisi, Muller, & Fitts, 1957) in indicating that average information transmission rates increase, within limits, as the size of the coding alphabet is increased. These results, it should be noted, are for relatively naive *S*s whereas most previous studies have employed practiced *S*s.

Various forms of linear equations were fitted to the data collected here, including the form proposed by Hick in 1952, $\text{RT} = b \log (n + 1)$, where n is the antilogarithm of H_t. The best fit was obtained for the form used in Table 1, $\text{RT} = a + b(H_t)$. This mathematical model offers the additional advantage of providing estimates of RT as H_t approaches zero.[5]

Information transmission rate does not appear to be differentially affected by changes from discrete to self-paced serial reactions. There was a tendency for the discrete reactions to be made more accurately but more slowly than the serial; however, these differences in H_t and RT were small and compensating. Thus, the linear equations were essentially identical for each of the four comparisons. It appears, therefore, that the serial task, in which only the .15-sec. machine lag intervened between *S*'s response and the appearance of the next stimulus, permitted as effective information processing as did a rest interval of 8 sec. followed by a 2-sec. warning signal.

Marked differences in the values of the intercept (*a*) and slope (*b*) constants were found as a function of the S-R coding combinations employed. In view of the number of previous studies that have obtained similar S-R compatibility effects, including studies comparing vocal and motor responses (Alluisi et al., 1957; Muller, 1955), it is clear that human information processing time is highly dependent on the manner in which stimuli and responses are coded as well as upon the degree of stimulus and response uncertainty. The importance of such compatibility effects is further indicated by results that show these effects to persist even after considerable training (e.g., Fitts & Seeger, 1953) and to become increasingly important as the size of the coding alphabet is increased (as shown here). There is also some evidence that the importance of compatibility effects may increase with age (Suci, Davidoff, & Surwillo, 1960; Welford, 1958), and several writers have proposed that the magnitude of these effects increases as the level of task stress increases.

The results obtained in the vocal-response-to-numerals task are basically different from the other findings. In both the discrete and the self-paced serial task, average RT was only slightly affected by the amount of information transmitted (H_t) per S-R event. As a matter of fact, RTs remained approximately constant for the 2- and the 3-bits/stimulus levels of stimulus uncertainty.

The failure of RT to decrease significantly as the number of alternative stimuli was reduced has been found (*a*) in studies that required *S* to read words with the number of possible stimulus words varying over a range from 2 to 256 (Pierce & Karlin, 1957; Sumby & Pollack, 1954), (*b*) in a study wherein *S* was trained for 45,000 trials on a task requiring finger responses to lights (Mowbray & Rhoades, 1959), and (*c*) in a study in which *S*s made finger responses to vibratory stimuli applied to the finger tips (Leonard, 1959). From a consideration of the tasks used in these studies, the common finding may be given a plausible explanation in terms of the extended learning involved. It seems reasonable to expect that the stronger the habit strength of the S-R combinations required in a specific task, the less steep should be the function relating RT to H_t.

However, in the present study, there

---

[5] Simple RTs of 5 *S*s (1 male and 4 women) who were members of the laboratory staff were measured with the same apparatus used in this study. Means of 0.316 and 0.440 sec. were obtained for the key pressing and vocal responses, respectively, to the numerals; means of 0.207 and 0.311 sec. were obtained for the same responses to the lights. The σ_m's of these means were 0.008, 0.022, 0.001, and 0.010 sec., respectively. Each mean lies between the discrete and the self-paced serial-reaction intercepts of the appropriate equations of Table 1. The relatively high absolute values for the RTs to numerals may reflect the rise times and lower intensities of the projected numerals; vocal RTs include a slight lag time of the voice key.

is little a priori basis for assuming stronger habit strength for numeral-vocal responses than for light-finger responses, i.e., for predicting less effect of H_t per response for the former type of coding. It appears possible, as an alternative, that the effects of learning may be reflected in the ability to deal effectively with particular sets of stimuli in combination. In the present case the 4–7 and 3–4–7–8 combinations of numerals may have been so incongruent with *S*'s previous experience in dealing with subsets of numerals that he responded as though the entire alphabet of 10 Arabic numerals was being presented. Such a cognitive set presumably would not operate in the case of the lights, since there is no a priori reason to expect a set of any particular size.

SUMMARY

This study investigated the effects on information processing of (*a*) the use of self-paced serial reactions vs. discrete reactions, and (*b*) the use of different stimulus and response codes (both numerals and lights were used as stimuli, and both vocalizations and finger movements were used as responses), and (*c*) the use of three levels of stimulus uncertainty (ranging from 1 to 3 bits/stimulus in unit steps). These several conditions, and the use of naive *S*s, provided further tests of the generality of the function, $RT = a + b(H_t)$.

The results were: (*a*) Reaction time (RT) was an increasing linear function of the average amount of information transmitted (H_t) per stimulus-response event for three of the four S-R pairings employed. The RTs obtained for vocal responses to Arabic numerals, however, were affected only slightly by number of alternatives in the range from 2 to 8. (*b*) The self-paced and discrete tasks gave very similar results. There was a slight tendency for the discrete reactions to be made more accurately but more slowly than the self-paced serial reactions, but the differences both in RT and H_t were small and compensating. The linear functions relating RT with H_t were essentially identical for the two modes of stimulus presentation.

These findings are interpreted as indicating the importance of overlearning in determining S-R compatibility effects. The findings also suggest that *S*'s familiarity in dealing with specific subsets drawn from familiar alphabets may also affect his information-handling rates, when restricted subsets of stimuli and responses are used.

REFERENCES

ALLUISI, E. A., MULLER, P. F., JR., & FITTS, P. M. An information analysis of verbal and motor responses in a forced-paced serial task. *J. exp. Psychol.*, 1957, **53**, 153–158.

ARCHER, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

CROSSMAN, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice-response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

FITTS, P. M., & SEEGER, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

GREGG, L. W. The effect of stimulus complexity on discrimination responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

LEONARD, J. A. Tactual choice reactions: I. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

MOWBRAY, G. H., & RHOADES, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

MULLER, P. F., JR. Verbalization as a factor in verbal versus motor responses to visual stimuli. Unpublished doctoral dissertation, Ohio State University, 1955.

PIERCE, J. R., & KARLIN, J. E. Reading rates and the information rate of a human channel. *Bell Sys. tech. J.*, 1957, **36**, 497–516.

SHANNON, C., & WEAVER, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.

SUCI, G. J., DAVIDOFF, M. D., & SURWILLO, W. W. Reaction time as a function of stimulus information and age. *J. exp. Psychol.*, 1960, **60**, 242–244.

SUMBY, W. H., & POLLACK, I. Short-time processing of information. *USAF HFORL tech. Rep.*, 1954, No. 54–6.

WELFORD, A. T. *Ageing and human skill.* London: Oxford Univer. Press, 1958.

(Received November 17, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 111–114

THE ABILITY TO FILTER NOISE FROM A VISUAL TASK WHEN THE NOISE AND SIGNAL ARE PRESENTED SEQUENTIALLY[1]

DWIGHT E. ERLICK

Aerospace Medical Laboratory, Wright-Patterson Air Force Base

Encoding has proved useful in the processing of excessive information due either to overload conditions or to irrelevance of some stimuli in a situation. In some information processing problems where the information can be categorized, the human processor may handle an overload by recoding the input into another code that contains fewer units or chunks, with more bits per chunk (Miller, 1956). However, when this is not feasible, one may resort to processing only a selective portion of the informational input, i.e., acting as a human information filter. This process of filtering might be viewed as an encoding problem whereby the unused information is encoded into a general category such as noise. Several investigations dealing with irrelevant stimuli might be interpreted in this manner, i.e., the observer acting as a noise filter. The general trend of these studies indicates that simpler tasks tend to be less influenced by noise (Archer, 1954; Green, McGill, & Jenkins, 1953); performance tends to decrease as the quantity of filtering required increases (Archer, Bourne, & Brown, 1955; Gregg, 1954; Hodge, 1959); performance tends to decrease when stimuli to be encoded as noise, i.e., filtered out, have previously been encoded in terms of relevant dimensions of the task (Hodge, 1959); the ability to filter out noise increases as a function of the ease of differentiating or categorizing the relevant from the irrelevant aspects of the task; and practice tends to increase the ability to filter out stimuli encoded as noise (Hodge, 1959; Morin, Forrin, & Archer, 1961). In all the foregoing investigations, the relevant stimuli (signal) and the irrelevant stimuli (noise) were presented in the same modality and occurred in spatial and temporal contiguity.

The purpose of the present experiment was to determine the effect of presenting the signal and noise in the same modality but temporally separated, i.e., sequentially. It might be hypothesized that with a temporal separation of signal and noise, *S* would *have* to attend to each and every unit of noise as it appeared. This procedure might be expected to degrade performance more than when the signal and noise are presented at the same time, a situation where it is more feasible to disregard larger units or chunks of the noise per unit time. It might also be hypothesized that an increase in noise would tend to decrease performance. Based on these hypotheses, a task requiring the sequential analysis of stimuli was selected. The task involved was the judgment of the relative frequency of two random sequential categories. Erlick (1962) has shown that the information transmitted, based on correct identifications of the more frequent category, is linear as a function of the frequency difference between the categories. The task also showed no significant effects due to practice. The major variable investigated in the following experiment was the effect on this temporal task of varying the amount of noise (irrelevant stimuli) presented in sequential relationship to the task.

---

[1] The data were collected at the Engineering Psychology Research Project, Antioch College.

METHOD

A punched paper tape reader was used to present a preprogramed random series of different letters which were displayed one at a time in the same location on a 3 × 4 in. screen by means of a rear projection system. The symbols, 3¾ in. high, were presented at the rate of 4 per sec. and were on for approximately half of each cycle, producing an equal on-off ratio. This rate and ratio prevented *S* from counting the individual symbols, yet produced a distinct off-time when the same symbol was repeated in sequence. The letters *A* and *B* were the two mutually exclusive categories, i.e., the signal. The 10 letters, *C, D, E, F, G, H, K, L, M, N,* were the irrelevant stimuli, i.e., the noise.

The ratios between the two categories were used: 24:20, and 30:20. These were selected to test the effects of different signal intensities and will be referred to as S_1 and S_2, respectively. On the basis of previous studies (Erlick, 1962) S_1 and S_2 would be predicted to transmit .15 and .49 bits of information, respectively, in terms of identifying the more frequently occurring category.

With S_1 having a total of 44 relevant letters and S_2 having 50 relevant letters, the signal-to-noise ratios as computed in terms of the number of relevant to irrelevant letters were: 1, 2, 4.9, 11, and 22 for S_1; and 1, 2, 5, 10, and 16.7 for S_2. A no-noise condition was also used for both S_1 and S_2.

The interspersing of the noise increases the time period between signal letters. To control for possible interactions between the effects of the noise and the variations in time, the equipment was designed so that the irrelevant letters could be switched off and blank spaces, i.e., empty time intervals, substituted for these letters. In effect, the equipment filtered out the noise, allowing for comparison of performance of the task under conditions with and without noise. Therefore, the interference with the signal was subdivided into the noise conditions (N) and the blank interval conditions (B). A trial required one judgment and consisted of observing a random series of one of the conditions BS_1, NS_1, BS_2, or NS_2 at one of the signal-to-interference ratios or the no-interference condition, which is the same as the no-noise and no-blank interval condition. A session consisted of 16 trials. An experimental period consisted of six sessions each with a different signal-to-interference ratio. The ratios within each of the six sessions were approximately the same, i.e., 1, 2, 4.9; and 5, 10; and 11, 16.7; and 22; and no-interference. All sessions except the no-interference condition were subdivided into a unit of eight B trials and a unit of eight N trials. Within each of these B and N units, S_1 and S_2 were each replicated four times. A separate random selection (with replacement) of the irrelevant letters was made from the 10 irrelevant letters for each of the 4 replicated signal-to-noise ratios within each N unit. Different random sequences of signal and noise letters were also used for each of the four replications. The sequence of S_1 and S_2 within each B and N unit was randomly determined, with a different random sequence being used for each session and each experimental period.

Two practice trials preceded each of the units within each session. Prior to the two practice trials, *E* switched the equipment to either Cond. B or N. The six sessions were presented to six groups of 16 *S*s each, according to a latin square design. The six groups were subdivided so that half of the *S*s received Cond. B first for all sessions within a period and the other half of the *S*s received Cond. N first for all sessions. All 96 *S*s had two experimental periods separated by at least 1 day, with the order of B and N being reversed for the second period. Thus, all *S*s made 8 observations at each of the different experimental conditions for a total of 192 observations. Each experimental period was punched on one continuous tape with an automatic stop programed between trials. The foregoing different randomizations were achieved to some extent by reversing the tapes for half *S*s, thus producing different sequential orders of the *A*s, *B*s, and signal-to-interference ratio. The ratios S_1 and S_2 were counterbalanced so that the letters *A* and *B* each appeared more frequently half the time.

The *S*s were instructed that other letters would appear in some of the series, but their task was to determine only which of the two letters *A* or *B* had appeared more frequently. The same general procedure was used as in previous experiments (Erlick, 1961), including running groups of paid college *S*s and offering a monetary bonus as a function of the number of correct answers.

RESULTS

The following analysis is based on the summated responses of all 96 *S*s. The percentage of correct identifications for each of the two signal intensities (S_1 and S_2) at each of the five signal-to-interference ratios for Cond. B and N and the no-interference condition, were converted into information transmitted. A graphic analysis based on this measure indicated no systematic interactions between preference for the relevant symbols *A* or *B* and any of the other experimental conditions. Also, the amount of information transmitted for the first and second experimental periods was the same, .27 bits, indicating no effects due to practice. Therefore, responses to both sets of symbols and both experimental periods were combined for the following analyses.

The information transmitted at the no-interference condition for signal intensities of S_1 and S_2 was .17 bits and .47 bits, respectively. Comparison of the no-interference condition and the interference conditions averaged over all signal-to-interference ratios for S_1 and S_2 showed the same information transmitted for both conditions at S_2 and a slight decrease (.04 bits) for the interference condition at S_1. The amount of information transmitted averaged over both signal levels and all signal-to-interference ratios was .26 bits for the noise condition and .28 bits for the blank-interval condition. Graphic analysis of the interactions between the noise and blank-interval conditions and signal-to-interference ratios for both signal intensities showed no systematic effects indicating no differences between any of the foregoing conditions (see Fig. 1).

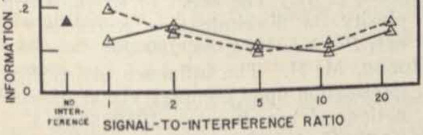


FIG. 1. Information transmitted as a function of the signal-to-interference ratio for noise and blank-interval conditions and for two different signal intensities. (The abscissa is scaled logarithmically. Each point is based on 192 observations except the no-interference condition which is based on 384 observations.)

DISCUSSION

The lack of any systematic bias by the group for the symbols *A* and *B* confirms the results of earlier studies by the author (Erlick, 1961, 1962). The fact that *S*s did as well when the signal-to-interference ratio was 1 as under the no-interference condition and also showed no significant differences between the noise and blank-interval conditions might be due to the nature of the noise used. Using easily discriminable, highly overlearned symbols to represent both the signal and noise probably enhances *S*'s ability to differentiate and put the symbols into categories of signal and noise (Bruner, Goodnow, & Austin, 1956; Dollard & Miller, 1950). The simplicity of the task might also be a factor in the negative findings.

Since the results were negative for the effects of degree of noise, one would not expect any improvement due to practice, and the results confirm this. The lack of any significant effects due to the sequential interference of noise supports the hypothesis that a temporal separation of signal and noise does not effect signal detection, at least for the type of noise used in this study. Similar results were obtained in a study when a signal and noise were simultaneously presented (Green et al., 1953); the color of a number to be searched for was specified in advance, and it was found that the presence of numbers of a clearly different color did not affect the search time. Archer (1954) used a visual task involving the identification of geometric patterns confounded with simultaneous noise and also found negative results in terms of the interference of the noise with the signal.

The predictions made on the basis of previous investigations by the author (Erlick, 1962) as to the amount of information that would be transmitted at the signal intensities of S_1 and S_2 were very accurate (±.02 bits). This lends validity to the proposed functional relationship between information transmitted in terms of identifying the more frequent category and the relative increment in frequency between categories. The fact that noise did not interfere with performance at either the S_1 or S_2 signal levels suggests that this relationship is constant for varying degrees of task difficulty.

Based on previous research by the author (Erlick, 1961), an improvement in perceiving the more frequent category of a random sequential series would be predicted as a function of decreasing the rate of presentation of the categories. It might be argued that the interspersing of blank intervals between the signal is equivalent to reducing the rate of presentation, i.e., a signal-to-interference ratio of 1 in this study reduces the average rate of signal presentation from 4 per sec. to 2 per sec. Consequently, an improvement in performance would be hypothesized as the signal-to-blank interval ratio increased. That no improvement in performance was found, might be explained in terms of the intervals being randomly interspersed throughout the signal instead of appearing as longer regular intervals as in the investigation of effects of rate. The lack of any significant differences between the noise conditions and the blank-interval conditions appears to indicate that for this task, the human could filter noise as effectively as the equipment.

SUMMARY

The ability of human *S*s to filter out noise from a task when the signal and noise are presented sequentially was investigated. Two levels of signal intensity and five signal-to-noise ratios (ranging from 1 to 22) plus a no-noise condition were used. A condition using blank time intervals instead of noise at all signal-to-noise ratios was also used. The task involved the judgment of the relative frequency of two random sequential categories (letters *A* and *B* presented visually). Ten different letters randomly interspersed among *A*s and *B*s represented the noise. There were no significant differences among any of the foregoing conditions, indicating that *S*s effectively filtered out the noise when it was presented in sequential relationship to the signal.

REFERENCES

ARCHER, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

ERLICK, D. E. Judgments of the relative frequency of a sequential series of two events. *J. exp. Psychol.*, 1961, **62**, 105–112.

ERLICK, D. E. Perception of the most frequent category of a random series as a function of the number of categories. *J. exp. Psychol.*, 1962, **63**, 115–118.

GREEN, B. F., MCGILL, W. J., & JENKINS, H. M. The time required to search for numbers on large visual displays. *MIT Lincoln Lab. tech. Rep.*, 1953, No. 36.

GREGG, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

HODGE, M. H. The influence of irrelevant information upon complex visual discrimination. *J. exp. Psychol.*, 1959, **57**, 1–5.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MORIN, R. E., FORRIN, B., & ARCHER, W. Information processing behavior: The role of irrelevant stimulus information. *J. exp. Psychol.*, 1961, **61**, 89–96.

(Early publication received August 31, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 115–118

PERCEPTION OF THE MOST FREQUENT CATEGORY OF A RANDOM SERIES AS A FUNCTION OF THE NUMBER OF CATEGORIES[1]

DWIGHT E. ERLICK

Aerospace Medical Laboratory, Wright-Patterson Air Force Base

An earlier paper by the author (Erlick, 1961) presented findings which indicated that perception of the more frequent category of a sequence of two different categories is a function of the sequence of the categories, the duration of observation, and the rate of presentation. When both categories are equally frequent, the category having the higher degree of clustering is estimated as having a higher relative frequency of occurrence. Also, when both categories are presented randomly, and one occurs more frequently, the percentage of the correct identifications of the more frequent category increases as a function of an increasing difference in the relative frequencies between the categories. The ability to identify the more frequent category was found to decrease as a function of increasing rate of presentation (from one to eight events per sec.) and decreasing duration of observation (from 40 to 10 sec.). The foregoing investigation used only two categories. In terms of information analysis (Shannon & Weaver, 1949), the uncertainty as to which category had occurred more frequently was equivalent to 1 bit.

Many investigations, as reviewed by Miller (1956), have been concerned with the relation of amount of information input to information transmission. The findings generally indicate that as information input increases there is an initial increase in information transmission, then a leveling off at what is called channel capacity, and finally a decrease in transmitted information as overloading takes place.

The purpose of the present investigation was to increase the range of information inputs and determine two relationships: that between information transmitted and the frequency difference between categories; and that between information transmitted and information input.

METHOD

The same general procedure was used as in previous experiments (Erlick, 1961). A punched paper tape reader presented a preprogramed series of the mutually exclusive categories (symbols) on a single plane, in-line display unit. The unit contained 12 lenses, each etched with one of the numerical symbols. By turning on and off individual lights behind each of the lenses, the symbols were projected one at a time, so as to be centered in the same location of a 3 × 4 in. frosted plastic surface. The symbols, which were 3¾ in. high, were seen as an outline of light against a dark background. A motor-driven cam supplied pulses to the tape reader which controlled the rate of presentation of the symbols. Each symbol was presented for approximately half of each cycle, producing an equal on-off ratio. Based on the author's previous work a rate of four symbols per sec. was selected. This rate was well beyond *S*s ability to count each symbol as it appeared, yet was sufficient so that a distinct off-time could be noticed when the same synbol was repeated in sequence. All of the programing equipment was housed in a separate room to reduce noise.

The number of different symbols presented during one trial varied from 2 to 12 in four steps: viz., 2, 4, 8, or 12, equivalent to 1, 2, 3, and 3.6 bits of information input (H) respectively, and hereafter referred to as the H conditions. The numerical symbols were sets of consecutive integers beginning with 1, e.g., in the four-symbol condition, 1, 2, 3, and 4 were used. Within each condition a basic frequency of 20 was used for all except one symbol, which had an increased frequency of one of five values. Categories (symbols) having frequencies of 20 are designated by C_L, while C_M is used to represent the frequency of that category with the increment. The percentage increments between C_M and C_L are designated by ΔF in the formula:

---

[1] The data were collected at the Engineering Psychology Research Project, Antioch College.

$$\Delta F = \frac{C_M - C_L}{C_L} \times 100$$

Given a constant ΔF, it was hypothesized that as H increased, the accuracy of perceiving C_M would decrease. Therefore, in order to obtain somewhat equivalent accuracy in responses to the different H conditions, the ΔFs were increased as H increased (see Table 1).

TABLE 1

PERCENTAGE INCREMENTS (ΔF) OF THE MORE FREQUENT CATEGORY (C_M), PERCENTAGE OF CORRECT IDENTIFICATIONS OF C_M, AND INFORMATION TRANSMITTED (T), AS A FUNCTION OF INFORMATION INPUT (H)

H	ΔF	% Correct C_M Identifications	T
1	5	55.38	.01
	10	59.38	.03
	15	62.50	.05
	30	80.03	.28
	50	87.50	.46
2	10	30.90	.01
	20	35.42	.04
	30	50.87	.22
	50	64.93	.51
	90	82.99	1.07
3	15	20.66	.04
	30	32.64	.20
	50	37.85	.30
	100	63.89	1.04
	180	86.11	2.03
3.6	15	12.67	.02
	30	18.58	.08
	50	27.95	.24
	125	58.85	1.18
	210	79.69	2.15

A trial required one judgment and consisted of *S* observing a random series of either 2, 4, 8, or 12 different categories at one of the ΔF values. A session was restricted to the same H condition and consisted of three replications of each of five ΔFs randomly presented. An experimental period consisted of four sessions, each with one of the four different H conditions. The four H conditions were presented to four groups of 24 *S*s each. A latin square design was used to counterbalance order. For each H condition all the categories were counterbalanced over sessions and groups so that each occurred as the more frequent (C_M) category the same number of times. At the end of each trial, *S* was instructed to write the number which appeared most frequently. The *S*s were told to guess if they were not sure of the answer and to keep all previous answers covered with a card with which they were provided. All 96 *S*s had two experimental periods separated by at least 1 day, and made a total of 30 judgments of the four H conditions. Sessions were punched on one continuous tape with an automatic stop programed between trials. The *E*, who sat in the room with *S*s, called out the number of each trial before it was presented and started the trial by pressing a button. The *S*s were run in groups of up to 10 at one time, with each *S* seated in an individual booth facing the display and viewing the symbols from a distance of from 6 to 9 ft. The *S*s were given three practice trials before each session. An attempt was made to motivate *S*s by offering a monetary bonus as a function of the number of correct judgments.

Results

The following analysis was based on the summated responses of all 96 *S*s. The percentages of correct identifications for each of the five frequency differences (ΔF), at each of the four levels of information input (H) were converted into information transmitted (T), as seen in Table 1. This measure (T) is derived by assuming that all categories within an H condition have equal likelihood of being selected as C_M. Thus, for example, in terms of information analysis (Shannon & Weaver, 1949), 75% identification of one of two equally likely events (H = 1) trans-

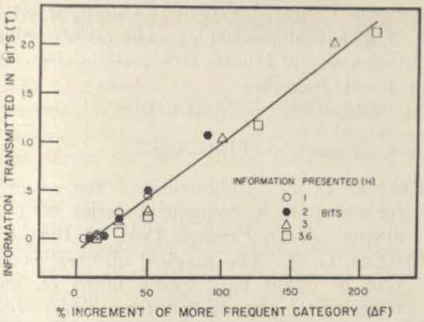


FIG. 1. Information transmitted as a function of the increment of the more frequent category for different information inputs. (Each point represents 576 observations.)

mits .19 bits, while 75% identification of one of eight equally likely events (H = 3) transmits 1.49 bits of information.

Graphic analysis indicated: no systematic interaction between symbol preference and the four H conditions; also no systematic effects for any of the experimental conditions from Session 1 to Session 2. Thus, the responses to trials in which different symbols appeared as C_M (for each ΔF within each H condition) were pooled over both sessions in the following analyses.

An examination of the relationship between T and ΔF (see Fig. 1) for each of the four H conditions of 1, 2, 3, and 3.6 bits gave the following equations based on the method of least squares with the subscript on T indicating H: $T_1 = .0107\ \Delta F - .0726$; $T_2 = .0138\ \Delta F - .1825$; $T_3 = .0123\ \Delta F - .2028$; $T_{3.6} = .0113\ \Delta F - .2427$. The overall equation based on all four H conditions is: $T = .0113\ \Delta F - .1342$, with the standard error of estimate, $\sigma_{T.\Delta F} = .09$.

Further analysis of the relationship between T and ΔF as indicated by the slopes of the equations shows approximately the same linear function for all H conditions. Using the slope of the overall equation, .0113, an increment of approximately 88% in frequency of a category is required to transmit 1 additional bit of information in terms of identifying a category as having occurred more frequently. Analysis of the relationship between T and H as indicated by the intercepts of the equations shows a slight decrease in T as H increases.

DISCUSSION

The lack of any systematic effects from Session 1 to Session 2 indicates no practice effects. The equations show approximately the same slope between T and ΔF for all four H conditions (1, 2, 3, and 3.6 bits) and there is a small standard error of estimate of the overall equation derived from all H conditions. Therefore, the ability to perceive the most frequently occurring category in a random series can be expressed as a linear relationship between reduction of uncertainty in identifying the more frequent category and the frequency differences between categories, at least for a range from 2 to 12 categories.

The lack of perfect information transmission even at the higher frequency increments is probably due in part to errors in recording by *S*s. Current investigations by the author involving absolute estimation of proportions show a wide disparity between actual and estimated proportions for some of the judgments, e.g., an actual proportion of 85% is occasionally estimated as 10%.

The slight decrease in T as H increased indicates that an overload condition possibly exists at all H conditions. This is not surprising in view of the fast rates of presentation of the categories.

Using the slope (.0113) between T and ΔF based on all H conditions and the intercept for the H_1 condition, the result of a previous investigation (Erlick, 1961) involving the same variables but different *S*s is predicted within ±.05T. This finding adds validity to the foregoing relationships between T, ΔF, and H.

SUMMARY

The ability to perceive the more frequently occurring category in a random series as a function of the number of different categories was investigated. The main variables of concern were: the relationship between information transmitted (T) and (*a*) the frequency difference between categories (ΔF), and (*b*) the information input (H). The different categories were 2, 4, 8, or 12 (H in bits = 1, 2, 3, and 3.6, respectively). Within each H condition a basic frequency of 20 was used for all except one category, which had an increased frequency of one of five values. Categories were presented randomly at the rate of four per sec. The *S*s were instructed to indicate the category which appeared most frequently. The results indicate that approximately the same linear relationship exists between T and ΔF for all H conditions with $T = .0113\ \Delta F - .1342$. The relationship between T and H indicates a slight decrease in T as H increases.

REFERENCES

ERLICK, D. E. Judgments of the relative frequency of a sequential series of two events. *J. exp. Psychol.*, 1961, **62**, 105–112.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

SHANNON, C. E., & WEAVER, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.

(Early publication received August 14, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 119–123

MANIFEST ANXIETY SCALE SCORE AND THE READY SIGNAL IN CLASSICAL CONDITIONING[1]

WILLIAM F. PROKASY, JR. AND FRANCIS L. WHALEY

Pennsylvania State University

With the construction of the Taylor Manifest Anxiety scale (Taylor, 1953), a number of studies (Baron & Connor, 1960; Caldwell & Cromwell, 1959; Spence & Farber, 1953; Spence, Farber, & Taylor, 1954; Spence & Taylor, 1951, 1953; Spence & Weyant 1960; Taylor, 1951) have reported a positive relationship between conditionability and score on the *MA* scale (Manifest Anxiety scale). Similarly, Spence and Beecroft (1954) and Spence and Farber (1954) report greater responsiveness to the reinforced stimulus in the classical discrimination learning situation, as well as some trend (nonsignificant) for better discrimination between reinforced and nonreinforced stimuli, in *S*s who scored high, rather than low, on the *MA* scale.

A problem is posed, however, by the studies reporting negative, or reverse findings. Hilgard, Jones, and Kaplan (1951) found greater classical discrimination differences in low-*MA* *S*s, rather than in high. Prokasy and Truax (1959) reported no difference between high- and low-*MA* score groups, while King (1958) found low-*MA* score *S*s averaging significantly greater responding than did high-*MA* score *S*s.

It is the King (1958) study which provides the clue for an explanation of these discrepancies. In a standard eyelid conditioning situation, half of the *S*s were high-*MA* scorers and the other half were low-*MA* scorers. Each of these groups was subdivided into two subgroups: one subgroup received a "ready, blink" signal 2 to 4 sec. before each conditioning trial, and the other subgroup did not receive this warning signal. In addition to the surprising (and to date unexplained) observation that in the presence of the ready signal the low-*MA* *S*s were the more frequent responders, it was found that there were no differences between high- and low-*MA* *S*s without the ready signal preceding each trial.

Thus, studies using a ready, blink signal have uniformly found that high-*MA* *S*s respond more frequently and tend to discriminate better than low-*MA* *S*s. When a blink was not required (Hilgard et al., 1951), the latter condition did not obtain. When a ready signal was not employed, the results are conflicting: Baron and Connor (1960) and Spence and Weyant (1960) report positive results, while Prokasy and Truax (1959) and King (1958) report negative results.

It is the purpose of this study to explore further the role of the ready, blink signal as a potential interacting factor in the high- vs. low-*MA* score conditioning studies reported earlier, with present emphasis placed upon classical discrimination learning.

METHOD

Apparatus.—The eyeblink conditioning apparatus has been described in detail elsewhere (Prokasy & Whaley, 1961). The *S*s were seated from 30 to 40 in. in front of a panel including a ½-in. orange fixation light (445 ml.) mounted midway between two milkwhite disks of 3-in. diameters. The disks were mounted vertically, their centers being separated by 8 in. The reinforced and unreinforced CSs were increases from 3.8 to 190 ml. illumination of the upper or lower disk. For half of the *S*s the upper disk was reinforced, and for half the lower disk was reinforced. The UCS was 1.5 liter/min puff of nitrogen presented through a 5-mm. (inside diameter) glass jet situated within .25 in. of *S*'s eyelash. Trials were automatized by a Western Union teletype transmitter with a fixed intertrial interval of 15 sec. Hunter-Brown interval timers controlled the interstimulus interval of 500 msec., the CS duration of 500 msec., and the UCS duration of 50 msec. The CS and UCS signals, the response, and the first time derivative of the response were recorded by a capillary pen recording system on a paper drive operated at 150 mm. per sec.

---

[1] This study was supported by funds from National Institute Mental Health Grant M-2560A and from National Science Foundation Grant 7463 to the senior author. The authors wish to acknowledge the technical assistance provided by Burton Klinger and Patricia Fitzgerald.

Subjects.—At the beginning of the semester all students in two large introductory classes at Pennsylvania State University were given the *MA* scale inventory under the guise of a "reliability" study. Those students scoring at and below 9 or at and above 21 were later requested to serve as *S*s in an experiment, but were told that they had been selected from the class rolls "at random." The means and *SD*s, respectively, of the two *MA* score groups are: High-*MA* score group, 26.19, 4.63; Low-*MA* score group, 6.7, 1.77. A total of 160 *S*s participated in this experiment, and, of these, 8 were lost because of apparatus failure or *E* error.

Procedure.—The two independent variables are the presence or absence of a ready, blink signal before each trial and two ranges of *MA* scores. As *S*s came to the laboratory they were assigned alternately to a signal or a no-signal condition. To avoid any bias in the situation, *E* did not have, at the time of running *S*s, information on whether they fell into the high- or low-*MA* score groups. As a result, the numbers of *S*s in the signal and no-signal partitions were equal, but they were not equal within *MA* conditions. Thus, there were four groups of *S*s: High *MA*, ready signal with 44 *S*s; Low *MA*, ready signal with 32 *S*s; High *MA*, no ready signal with 38 *S*s; and Low *MA*, no ready signal with 38 *S*s. Of the 152 *S*s remaining, 15 met the criterion defined by Spence and Taylor (1951) as "voluntary responders" (i.e., half or more of their CRs fell in the latency range 200 to 300 msec.), and were not employed in the analyses. Seven voluntary responders were in the no-signal groups (5 high *MA* and 2 low *MA*) and 8 were in the signal groups (3 high *MA* and 5 low *MA*). In order to restore cell proportionality for statistical analyses, the data of 4 *S*s from the high-*MA*, ready-signal group and 3 *S*s from the low-*MA*, no-ready-signal group were randomly eliminated. This left a total of 60 *S*s in the low *MA* partition (30 ready signal and 30 no ready signal) and 70 in the high *MA* partition (35 ready signal and 35 no ready signal). All records were scored by a paid undergraduate assistant who did not know the nature of the study.

The *S*s were given a total of 90 trials, 45 to the reinforced CS and 45 to the nonreinforced CS, with positive and negative stimuli presented in an unsystematic order. Those *S*s in the ready, blink signal condition received this signal from 2 to 4 sec. prior to CS onset. The *S*s were instructed to keep the fixation light in view at all times and to maintain a passive attitude. After the session each *S* was cautioned not to discuss the experiment with other students, although

TABLE 1

MEANS AND *SD*s OF RESPONSE FREQUENCY TO LAST 20 REINFORCED AND LAST 20 NONREINFORCED STIMULI

MA	Signal				No Signal			
	Reinforced		Nonreinforced		Reinforced		Nonreinforced	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
High	10.89	5.71	7.12	4.82	12.03	4.06	8.89	4.54
Low	7.90	4.87	4.50	4.11	11.30	4.38	8.13	4.15

none was aware of the true purpose of the experiment.

RESULTS

Mean numbers of responses to the reinforced and nonreinforced stimuli over the last 40 trials (20 reinforced and 20 nonreinforced) for all four groups are presented in Table 1. Employing the difference between numbers of responses to the reinforced and nonreinforced stimuli as the index of discrimination learning, an analysis of variance was performed on the discrimination learning data. Based upon an average within groups mean square of 22.34, *F* ratios as a function of the *MA* and signal variables, as well as their interaction, were all less than unity. A post hoc analysis based upon 36 *S*s who scored above 26 and 24 *S*s who scored below 6 on the *MA* scale resulted in no additional differentiation among treatment combinations.

Table 2 provides a summary of an analysis of variance of total number of responses to the reinforced stimulus on the last 20 reinforced trials. Overall responding in the high-*MA* group is significantly higher than in the low-*MA* group, and overall responding in the no-signal group is significantly greater than in the signal group. As suggested by the means in Table 1, high- and low-*MA* groups did not significantly differ in the no-signal condition ($F = .52$; $df = 1/63$) but did significantly differ in the signal condition ($F = 5.1$; $df = 1/63$).

TABLE 2

ANALYSIS OF VARIANCE OF THE NUMBER OF RESPONSES ON THE LAST 20 REINFORCED TRIALS

Source	*df*	*MS*	*F*
MA level	1	111.43	4.82*
Ready signal	1	155.11	6.71*
MA × Signal	1	41.15	1.78
Within groups	126	23.12	—

* $P < .05$.

DISCUSSION

The findings indicate that prior results with the *MS* variable may be dependent upon anticipatory effects induced by the ready signal. The data of three prior studies, however, are not consistent with this conclusion. The first is that of Hilgard, Jones, and Kaplan (1951), in which the high-*MA* score group manifested less discrimination between reinforced and nonreinforced stimuli than did the low-*MA* score group. While these investigators employed the ready signal, they did not require that *S* blink before CS onset as is customary in the studies from the Iowa laboratory and as was done in the present study. The possibility that a blink response between signal and CS onset interacts with the *MA* score conditionability relationship remains to be explored.

The second such study is that of Spence and Weyant (1960) who compared the conditioning performances of high- and low-*MA* score groups without preceding each trial with a ready signal. In accord with past studies from the Iowa laboratory (cited in the introduction), the high-*MA* score group manifested a higher response frequency. Since their finding without the ready signal differs from those of King (1958), Prokasy and Truax (1959), and the present study, it is, perhaps, worth pointing out that Spence and Weyant introduced an irrelevant discrimination task into the conditioning situation in order to mask the nature of the experiment. Between most conditioning trials, *S*s were required to press a telegraph key in response to an auditory stimulus. Here, again, the possibility of an interacting influence cannot be overlooked, as the effects upon conditioning of an irrelevant discrimination task are unknown.

Finally, Baron and Connor (1960) obtained a rho of .4 between response frequency and *MA* score in classical eyeblink conditioning, but did not employ a ready signal preceding each trial. These results are similar to those of Taylor (1951) and it is not clear why they differ so markedly from those of King (1958) and the present investigators.

In conjunction with the present research, the Baron and Connor (1960) study suggests that the *MA* scale may separate *S*s in two ways: distinguishing high from low conditioners and, in addition, separating those *S*s who are made more susceptible to conditioning by virtue of a verbally induced preparatory set. Whether or not the effects of either aspect of the *MA* scale becomes manifest, however, may well depend upon rather subtle influences in the conditioning situation: influences such as specific UCS intensity, interstimulus interval, intertrial interval, or, possibly *E-S* rapport.[2]

That overall depression of response frequency was obtained in the presence of the ready signal is consistent with the results of studies by King (1958) and by McAllister and McAllister (1960), but is not compatible with the position of Spence and Weyant (1960) who posit at least a small increment in drive to be initiated with the presentation of the ready signal. Such a drive increment would, at the very least, lead to equal or greater performance with the ready signal than without it.

In conclusion, the observations that the presence of the ready signal reduces the overall level of responding and interacts with the manifestation of *MA* related conditionability suggests that the ready signal be more fully explored before it is adopted as a standard procedure in the classical conditioning situation. While it is true that there is no one "real" conditioning procedure, it appears evident that the ready signal is not a necessity for the development of CRs. To add it as part of conditioning methodology would seem to require that the investigator have a reasonable understanding of its operation as an interacting independent variable.

SUMMARY

With high and low *MA* scores and presence or absence of a ready signal as the independent variables, 152 *S*s served in a classical discrimination learning situation. The principal findings were (*a*) that the ready signal depressed overall frequency of responding, and (*b*) that high-*MA* score *S*s responded significantly more than did low-*MA* score *S*s only under the signal condition.

REFERENCES

BARON, M. R., CONNOR, J. P. Eyelid conditioned responses with various levels of anxiety. *J. exp. Psychol.*, 1960, **60**, 310–313.

CALDWELL, D. F., CROMWELL, R. L. The relationship of manifest anxiety and electric shock to eyelid conditioning. *J. exp. Psychol.*, 1959, **57**, 348–349.

HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.

KING, M. S. Anxiety and eyelid conditioning: The effect of a ready signal. Unpublished master's thesis, Duke University, 1958.

MCALLISTER, W. R., & MCALLISTER, D. E. The influence of the ready signal and unpaired UCS presentations on eyelid conditioning. *J. exp. Psychol.*, 1960, **60**, 30–35.

PROKASY, W. F., & TRUAX, C. B. Reflex and conditioned responses as a function of manifest anxiety. *Amer. J. Psychol.*, 1959, **72**, 262–264.

PROKASY, W. F., & WHALEY, F. L. The intertrial interval variable in classical conditioning. *J. exp. Psychol.*, 1961, **62**, 560–564.

SPENCE, K. W., & BEECROFT, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, **48**, 399–403.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, 116–119.

---

[2] G. A. Kimble (personal communication) and M. R. Baron (personal communication) have indicated that *E-S* interaction may play a role in the *MA* conditionability relationship.

SPENCE, K. W., & FARBER, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127–134.

SPENCE, K. W., FARBER, I. E., & TAYLOR, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of UCS as determinants of amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SPENCE, K. W., & TAYLOR, J. A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.

SPENCE, K. W., & WEYANT, R. G. Conditioning performance of high- and low-anxious *S*s in the absence of a warning signal. *J. exp. Psychol.*, 1960, **60**, 146–149.

TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received October 28, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 124–128

A TEST OF TWO INTERPRETATIONS OF THE APPARENT SIZE EFFECTS IN A DISTORTED ROOM[1]

WILLIAM EPSTEIN

University of Kansas

If two objects of equal physical size are placed in the two rear windows of a monocular distorted room, they will appear to be unequal in size. One explanation of such size effects which is frequently cited is formulated in terms of an interaction of assumptions, apparent distance, and retinal size. This interpretation is often associated with the transactionalist school (Kilpatrick, 1952) and may be called the *neoempiristic* viewpoint.

An alternative explanation is possible which makes no mention of assumptions or apparent distance. Instead, the effects are seen to depend entirely on the relational properties of the proximal stimulus configuration. This interpretation may be designated as the *relational* viewpoint.[2] If apparent size is relationally determined, then it follows that when the relational stimulus patterns which are being compared are not identical, the perceived size will not be identical as well. In the distorted room the two rear windows are retinally equal because the size of the farther window has been increased in proportion to the added distance. However, the retinal sizes of the physically identical objects whose size is to be judged are not equal. The more distant object projects a much smaller image. Therefore, the proximal stimulus relationships in the two fields are different, and the object in the more distant window will appear to be smaller. This outcome is determined by the retinal size of the stimulus objects relative to the retinal size of their respective frameworks. If this is indeed the case, then the introduction of the factors of apparent distance and "assumed normalcy" becomes superfluous.

The purpose of the present investigations was twofold: (*a*) a quantitative estimate of the magnitude of the anomalous effects was sought, and (*b*) an attempt was made to determine whether relational factors would account for the apparent size effects.

EXPERIMENT I

The literature dealing with the distorted room (e.g., Ittelson & Kilpatrick, 1952; Kilpatrick, 1954; Wiener, 1956; Wittereich, 1959) reveals a common tendency to describe the obtained size distortions qualitatively. For this reason, Exp. I was designed to provide a quantitative estimate of the distortions of apparent size.

Method

Apparatus.—The main apparatus is pictured and described in detail elsewhere (Ittelson, 1952, pp. 40–43; Ittelson & Kilpatrick, 1952, pp. 41–55). It was the distorted room labeled "Monocular Distorted Room No. 1" in Ittelson's (1952, p. 41) manual. For the present purposes the following details are necessary. The back wall was tilted from the frontal parallel plane from right to left away from *S*. The distance from the reference point to the center of the base of the right window was 40.25 in., while the distance to the same point in the left window was 46.38 in. The mean heights of the right and left windows were 19.50 in. and 26.50 in., respectively.

---

[1] The work described in this paper was supported by a grant (M-4153) from the National Institute of Mental Health of the United States Public Health Service and a grant from the General Research fund of the University of Kansas. Thanks are due to Raymond Engstrand who collected the data and assisted in the statistical computations.

[2] See Rock and Ebenholtz (1959) for a recent discussion of the relational determination of apparent size.

Two rods, painted white, were placed in position at the objective centers of the bases of the two windows. The rod in the right window was 0.5 in. in diameter, while the rod in the left window was 0.625 in. in diameter. During each judgment the rod which served as the standard was kept at a constant height of 10 or 15 in. By turning a wheel *S* could produce changes in the height of the comparison rod.[3] At one extreme the rod could be eliminated from view, and at the other extreme it could be made to span the entire length of the window.

Subjects.—The *S*s in Exp. I were 40 students of introductory psychology. None of the *S*s wore corrective glasses and all indicated that they had normal vision.

Procedure.—Each *S* was assigned to one of the following four experimental conditions: Cond. 1a ($N = 10$): Standards on left, 15-in. standard first; Cond. 1b ($N = 10$): Standards on left, 10-in. standard first; Cond. 2a ($N = 10$): Standards on right, 15-in. standard first; and Cond. 2b ($N = 10$): Standards on right, 10-in. standard first.

Two ascending and two descending judgments of apparent equality for each standard were obtained by the equation method. The judgments for each standard were made successively in a balanced order. The comparison stimulus was set at an initial height of 2 in. from the bottom of the window for the ascending trials; and for the descending trials it was set at a height which was 2 in. from the top of the window. No time limit was placed on *S*'s setting, but a record was kept of the time required for each setting. The times ranged from 5 to 10 sec.

When *S*'s head was in the desired position he was asked to describe the room. Neither of the two rods was visible during this period. Then the instructions were read by *E*. The instructions encouraged *S* to base his judgments on immediate, uncalculated perceptual impressions of size.

TABLE 1

MEAN JUDGMENTS (IN.) OF APPARENT EQUALITY IN EXP. I AND II

Size of Standard	Condition 1		Condition 2	
	Mean	*SD*	Mean	*SD*
10 in.	7.31	0.24	13.74	0.29
15 in.	10.93	0.33	20.73	0.27
8 in.	6.22	0.56	10.22	1.28
12 in.	9.26	0.72	15.21	1.81

After the final judgment *S* was questioned about the appearance of the distorted room and also about the procedure for arriving at a satisfactory match.

Results

The mean judgments of apparent equality were nearly identical for Cond. 1a and 1b, and the same was true for Cond. 2a and 2b. For this reason the results for Cond. 1a and 1b have been combined, and the same has been done for Cond. 2a and 2b. A summary of the data is contained in the upper two rows of Table 1. Although these results were not unexpected, they are nonetheless quite striking. If apparent equality can be considered to be a transitive relationship, then the data for the 10-in. and 15-in. standards, respectively, may be expressed in the following way: 7.31 in. = 10.00 in. = 13.74 in., and 10.93 in. = 15.00 in. = 20.73 in.

All of the distributions were highly homogeneous as indicated by the *SD*s given in Table 1. Similar homogeneity was observed in *S*s' descriptions of the distorted room. All *S*s described it as a room whose appearance was completely normal. The postexperimental interview revealed that *S*s were unaware of the objective nature of the stimulus situation, as they were at the outset of the experiment.

---

[3] The measuring device was not visible to *S*, nor was there any opportunity for associating any specific position of the wheel with a particular size.

EXPERIMENT II

In order to determine whether the size effects were relationally determined, a stimulus situation was created in which the neoempiristic and relational viewpoints would predict opposed, mutually exclusive outcomes.

Method

Apparatus.—The apparatus was the same as in Exp. I with the exception of three modifications: (*a*) The major change consisted of reducing the size of the left, more distant window to the dimensions of the right, nearer window. (*b*) Both rods were 0.5 in. in diameter and were painted the same medium gray as the window frames. (*c*) The standard rods were 8 and 12 in. in height.

Subjects.—The *S*s were 48 students from the same population. All had normal, uncorrected vision.

Procedure.—The procedure was identical to that followed in Exp. I with the exception of two minor alterations. The following sentence was added to the instructions: "It is equally important that when you have completed your match the rods look to you to be equal in length with respect to each other, and not necessarily with respect to the windows."[4] The second difference in procedure was an addition to the post-experimental interview. The *S* was requested to "draw the room as you remember seeing it."

Under the modified stimulus conditions of this experiment the two interpretations require different outcomes. Assuming that the normal rectangular appearance of the room is unaffected by the alterations described above, then the neoempiristic theory predicts size matches comparable to those obtained in Exp. I. Relational considerations lead to a different prediction. The standard and comparison should appear to be equal when they are in fact objectively equal. This requirement follows from the observation that only when the two rods are objectively equal will the two retinal fields be identical relationally. The same considerations apply irrespective of the location of the standard and, therefore, the prediction holds both for Cond. 1 and Cond. 2.

Results

Eight *S*s showed by their pre-experimental comments and post-experimental drawings that they had perceived the true shape of the distorted room. Therefore, the data reported for this experiment are based on the responses of only 40 *S*s who were evenly distributed among the four experimental conditions. None of the 40 *S*s whose responses were accepted reported any perceived distortion. There was only one major difference in the appearance of the room between the two experiments. In Exp. I both windows were perceived as equal in size; in Exp. II the left window appeared to be considerably smaller than the right window.

The results for Cond. 1a and 1b have been combined, and the results of Cond. 2a and 2b have been combined. A summary of these data is contained in the bottom two rows of Table 1. The results were in close accord with the neoempiristic expections and are comparable to those obtained in Exp. I.

DISCUSSION

The outcome of Exp. I is compatible with both interpretations. However, the results of Exp. II are consonant only with the neoempiristic viewpoint. The neoempiristic thesis requires that the adjustments in the physical size of the comparison should be in direct proportion to the difference in radial physical distance between the rods. The requirements of size-distance proportionality for both experiments are contained in Table 2 together with the deviation values, i.e., the amount of discrepancy between the obtained (see Table 1)

---

[4] This segment of the instructions was adopted with only minor changes from Rock and Ebenholtz (1959, pp. 389–390). The original version refers to "lines" and "frames" instead of rods and windows.

TABLE 2

VALUES (IN.) REQUIRED BY NEOEMPIRISTIC INTERPRETATIONS AND PERCENTAGE DEVIATIONS FROM THE REQUIRED VALUES

Size of Standard	Condition 1		Condition 2	
	Required Value	% Deviation	Required Value	% Deviation
10 in.	8.50	−14	11.50	+19
15 in.	13.50	−19	17.50	+18
8 in.	7.00	−11	9.00	+14
12 in.	10.50	− 7	14.00	+ 8

Note.—% Deviation is $\frac{\text{Req. Value}}{\text{Deviation}}$. A negative sign indicates that the obtained match was short of the required height. A positive sign indicates that the match exceeded the required height.

and the required values. In Exp. I the magnitude of the deviations from the values required by the neoempiristic view ranged from −14% to +19%; in Exp. II the range was −11% to +14%.

There are several possible reasons for the ineffectivess of relational factors in the present situation: (*a*) Unlike the Rock and Ebenholtz (1959) studies, the rods and their immediate frameworks were not presented in isolation. Both stimulus fields were perceived simultaneously within a common frame of reference, i.e., the back wall. The simultaneous presence of several subordinate and superordinate frameworks may have vitiated the influence of the immediate framework. (*b*) Another reason concerns the logical organization of the experimental task. Under Cond. 1, *S* is shown a standard which spans a specified length of a phenomenally small window. On his right *S* sees a window which is subjectively much larger. The relational view requires that the comparison in the right window be made objectively equal to the standard to achieve apparent equality. In fact, the ideal *S* would adjust the size of the comparison rod so that it spanned the same proportion of the window as does the standard in its window. However, a setting of this nature means that two rods will be judged equal in size when they span equal proportions of *phenomenally differently sized windows.* This is not a logically tenable outcome. Nevertheless, it is the result predicted by the relational point of view. Thus, the relational effect is opposed by the necessity for internal consistency and, as a result, may have been overriden. However, it should be noted that the instructions in Exp. II were deliberately formulated in a manner which was intended to discourage intentional judgmental manipulation. The *S*s were asked to make an immediate judgment of apparent size, and very few indicated that they had done otherwise. The fact that all *S*s completed their judgments in 10 sec. or less is presumptive evidence against the intrusion of deliberate judgment.

The present investigation gives little reason to question the critical role of apparent distance in producing the anomalous size effects obtained in the distorted room. However, there is one reservation which should be mentioned with regard to the usual formulation of the neoempiristic thesis. It is hard to see the justification or necessity for introducing "assumptions," "predicdictions," or "involuntary bets" into the explanation. The results seem to be explicable entirely in terms of an interaction between retinal size and apparent distance.[6]

SUMMARY

The anomalous apparent size effects observed in a distorted room may be explained in two ways. The *neoempiristic* interpretation explains the effects in terms of an interaction of assumptions, apparent distance, and retinal size. The relationalist interpretation refers the effects to the relational properties of the proximal situation.

Two experiments concerned with the anomalous apparent size effects in a monocular distorted room were reported. Experiment I provided a quantitative estimate of the obtained size distortions. The data were compatible with both the neoempiristic and the relationalist viewpoints.

---

[6] See Linksz (1952, pp. 628–629) for a complete exposition of this point.

In Exp. II a stimulus situation was created which made possible a test of the two explanatory alternatives. The construction of the distorted room was modified in a manner as to lead the two interpretations to predict mutually opposed size distortions. The results fit the neoempiristic interpretations but were not consistent with an explanation in relational terms.

REFERENCES

ITTELSON, W. H. *The Ames demonstrations in perception.* Princeton: Princeton Univer. Press, 1952.

ITTELSON, W. H., & KILPATRICK, F. P. Equivalent configurations and the monocular and binocular distorted rooms. In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N. H.: Institute of Advanced Studies, 1952.

KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view.* Hanover, N. H.: Institute of Advanced Studies, 1952.

KILPATRICK, F. P. Two processes in perceptual learning. *J. exp. Psychol.*, 1954, **47**, 362–270.

LINKSZ, A. *Physiology of the eye.* Vol. 2. *Vision.* New York: Grune & Stratton, 1952.

ROCK, I., & EBENHOLTZ, S. The relational determination of perceived size. *Psychol. Rev.*, 1959, **66**, 387–401.

WIENER, M. Perceptual development in a distorted room: A phenomenological study. *Psychol. Monogr.*, 1956, **70**(16, Whole No. 423).

WITTEREICH, W. J. Visual perception and personality. *Scient. American*, 1959, **200**, 56–60.

(Received November 9, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 129–142

CONFUSION MATRIX ANALYSIS OF SINGLE AND MULTIDIMENSIONAL AUDITORY DISPLAYS[1]

MILTON H. HODGE

University of Georgia

AND IRWIN POLLACK

Operational Applications Office, Bedford, Massachusetts

A major goal in psychophysical experimentation is the prediction of the complete pattern of confusions among alternative stimulus objects and their associated responses. An important step in this direction, the *constant-ratio rule* (CRR), has been furnished by Clarke (1957). The rule has been successfully applied in a number of situations employing "complex" informational displays, e.g., words or syllables presented in noise (Clarke, 1957; Clarke & Anderson, 1957; Pollack & Decker, 1960) and words flashed briefly with a tachistoscope (Anderson, 1959). However, a preliminary experiment (Clarke, 1959) with elementary auditory displays, e.g., the frequency and intensity of a tone, has suggested that the rule may break down. Considering the subtle complexities associated with the responses to such displays, the failure of the CRR—or any other single rule—is reasonable. Consequently, the present experiments were undertaken to provide a better foundation for making a decision about the rule with respect to elementary displays.

Definitions

Before examining the rule and some of its implications, consider the following definitions:

Stimulus ensemble.—A well-defined set of stimulus objects; e.g., eight sound levels of a 1000-cps tone, 0.25 sec. in duration, equally spaced on a db. scale between sound levels of 81 and 88 db.

Response set.—A well-defined set of responses corresponding one to one with the objects of the stimulus ensemble; e.g., the cardinal numerals 1–8 corresponding with the ensemble of eight stimulus objects.

Stimulus subensemble.—An arbitrary selection of stimuli from a given stimulus ensemble; e.g., any four or any two of the eight sound levels in the stimulus ensemble.

Single-dimensional ensemble.—A stimulus ensemble in which the stimulus objects differ only with respect to a single variable; e.g., the intensity or the duration of a tone.

Multidimensional ensemble.—A stimulus ensemble in which the stimulus objects differ with respect to more than a single variable; e.g., the frequency, intensity, and duration of a tone. Another example is speech.

Identification.—An experimental task in which *S* attempts to match members of the response set with corresponding members of the stimulus ensemble; the task required in the method of absolute judgment.

Stimulus-response confusion matrix.—A description of an identification experiment in terms of the entire distribution of responses to each member of the stimulus ensemble; e.g., the matrix of Table 1 is a 4 × 4 confusion matrix.

Correct response.—An identification response which matches the corresponding member of the stimulus ensemble; a response on the negative diagonal of the confusion matrix; e.g., the bold face entries of Table 1.

Incorrect response.—A response which does not match the corresponding member of the stimulus ensemble; a response not on the negative diagonal.

Observed confusability.—The mean proportion of observed incorrect responses associated with a designated stimulus subensemble.

Derived confusability.—The mean proportion of calculated incorrect responses associated with a designated subensemble after the arbitrary removal of the other members of

---

[1] This is Technical Report 60-43 of the Air Force Command Control Division. The research is in support of Project 7682, Information Processing by the Human Operator.

TABLE 1

A STIMULUS-RESPONSE CONFUSION MATRIX

Stimuli	Responses			
	1	2	3	4
I	a	b	c	d
II	e	f	g	h
III	i	j	k	l
IV	m	n	o	p

Note.—Derived confusability scores:

$$(\mathrm{I, II}) = 0.5\left[\frac{b}{(a+b)} + \frac{e}{(e+f)}\right]$$

$$= \frac{b+e}{(a+b+e+f)} \quad \text{when } a+b=e+f$$

$$(\mathrm{I, IV}) = 0.5\left[\frac{d}{(a+d)} + \frac{m}{(m+p)}\right]$$

$$= \frac{d+m}{(a+d+m+p)} \quad \text{when } a+d=m+p.$$

Each alphabetic entry represents the conditional probability of a specific response (indicated by the column number), given the stimulus (indicated by the row number). When stimuli are equally likely, the conditional response probabilities are proportional to the original matrix entry.

the ensemble and their associated responses; e.g., algebraically defined in Table 1 for two particular pairs of stimulus objects.

Contextual effects.—Changes in the derived confusability of a given subensemble which result from a particular definition and selection of other members of the stimulus ensemble; e.g., changes in derived confusability occurring as the result of changes in the spacing among the stimulus objects.

Constant-ratio rule.—An empirical rule which states that the derived and observed confusability of any particular subensemble are identical and that both are independent of the other members of the stimulus ensemble; e.g., if a specific subensemble were embedded in ensembles of different size, the rule would predict that the derived confusability would be invariant with the size of the ensemble.

Implications of the Constant-Ratio Rule

The CRR implies the lack of stimulus and response interactions other than those between pairs of stimuli and their associated responses. The confusability of a given stimulus pair and its associated responses is assumed to be independent of other stimuli and responses in the matrix. That is, when stimuli and responses are eliminated from the confusion matrix, the CRR assumes that the conditional probabilities of the remaining members of the matrix do not change. For example, consider the ensemble of four stimuli I, II, III, and IV with associated Responses 1, 2, 3, and 4, illustrated in Table 1. The CRR suggests the following operations in obtaining the derived confusability between I and II and their responses. We first obtain the initial 4 × 4 matrix as shown in Table 1. We then dispose of Stimuli III and IV and of Responses 3 and 4. Finally, the remaining 2 × 2 matrix is normalized so that its conditional response probabilities sum to 1.0. The rule predicts that the resulting derived confusability will be identical with the observed confusability of a matrix consisting of Stimuli I and II and Responses 1 and 2.

The postulated lack of stimulus and response interactions is in direct opposition to numerous psychophysical studies demonstrating adaptation-level, anchoring, and response-bias effects with single dimensional stimuli. These effects all imply that the interactions are a function of the stimulus ensemble and response set in which the stimuli and responses are embedded. An extreme example of the interaction approach is the rubber-band model (RBM) of sensory organization (McGill, 1960). In a particular discrimination situation, e.g., identification of the objects of an ordered stimulus ensemble, the model assumes that *S* has a given discriminative capacity available for the discrimination task. If confusability is employed as a measure of *S*'s performance of the task, the RBM states that confusability is inversely proportional to discriminative capacity. Other *S*s performing the same task may have different confusability scores because of different discriminative capacities. The model further assumes that the discriminative capacity available for a given task is stretched over the range of the stimulus ensemble under examination. Schematically, imagine a rubber band with a fixed number of molecules (discriminative capacity), and capable of being stretched over a finite distance (the range of a stimulus ensemble). When stretched over a wide range, a lesser amount of the available discriminative capacity is assignable to a given stimulus separation (stimulus pair) than when stretched over a narrow range. Consequently, the confusability of two stimulus objects will increase with increases in the range of the ensemble. To summarize, the RBM assumes that the confusability of two stimulus objects is a function of the available discriminative capacity and the range of the stimulus ensemble. Confusability will increase with a decrease in discriminative capacity or with an increase in the ensemble range over which a given discriminative capacity must be stretched.

The CRR and the RBM thus represent opposite extremes of a continuum of postulated interaction, with the CRR assuming no interaction and the RBM assuming complete dependence upon the order and range of the stimulus ensemble. Although the assumptions of the RBM tend to restrict the model to single-dimensional ensembles, recent work with scaling problems (Torgerson, 1958) may make its application to multidimensional ensembles feasible. The CRR is not bound by the order and range assumptions, and is presumably applicable to both single and multidimensional ensembles.

The present experiments were designed to examine the accuracy of the CRR in predicting the performance of *S*s to a variety of auditory stimulus ensembles. More specifically, the experiments (*a*) evaluate contextual effects in single-dimensional ensembles, and (*b*) assess the adequacy of the CRR in both single and multidimensional ensembles.

GENERAL METHOD

In all, nine experiments were completed. The experiments are numbered in order of their execution, but for purposes of presentation, the separate experiments are reorganized.[2] Features common to the separate studies include:

Subjects.—A single crew of 7 university students served in all tests. They were tested in groups of 6, with each *S* enjoying a brief rest between successive experimental runs. All *S*s had extensive previous experience with psychophysical testing. The *S*s, reimbursed for their services, were employed 8 hr. a day over a period of 3 months. The first 10 testing days were devoted to preliminary practice in which asymptotic levels of performance were achieved.

Apparatus and procedure.—A stepping-switch programer selected among 2, 4, or 8 alternative stimuli. A single run was defined as 24 successive stimulus presentations; 48 different runs were available from the programer. The tests were carried out in groups of four runs. Each of the alternative stimuli occurred equally often in every run. The programer selected among eight durations, eight intensities, or eight combinations of frequency, intensity, and duration.

Each *S* was equipped with a response box which contained eight push buttons, eight corresponding lights, and a set of bulbs which marked the temporal sequence of Ready, Vote, Score. The responses of the 6 *S*s were summed directly into counters so that the complete stimulus-response matrix was available in real time. Only a single response button could be pressed at any one time. The correct response was indicated to *S*s at the end of each temporal sequence. In order to encourage individual performance, the number of correct responses per *S* was obtained and served as the basis of a monetary incentive system.

---

[2] Experiment V is not presented because an adequate amount of data is not available.

In each experiment *S*s were given extensive practice with the stimulus objects before data were recorded. Each study was completed before the next was begun.

Treatment of results.—For each stimulus-response matrix, the following quantities were calculated: (*a*) the observed conditional response probability per entry; (*b*) when applicable, the derived conditional response probability per entry; (*c*) the mean difference (without regard to sign) between the derived and observed probabilities; (*d*) the percentage of occurrences in which the derived and observed probabilities differ by a proportion greater than .10; and (*e*) the mean difference (with regard to sign) between the derived and observed probabilities for the diagonal matrix entries (the correct responses).

CONTEXTUAL EFFECTS IN SINGLE-DIMENSIONAL ENSEMBLES

Three experiments (III, IX, and VII) were specifically designed to examine contextual effects in single-dimensional ensembles. The absence of systematic contextual effects tends to support the CRR, whereas the presence of such effects tends to qualify the rule.

Experiment III: Ensemble Range

Experiment III was designed to study the derived confusability of a pair of stimulus objects as a function of the range of the stimulus ensemble.

Method.—A total of 21 stimulus ensembles, each consisting of four durations of a 1000-cps tone, were employed. The ensembles are schematically illustrated in Table 2. Each section of the table illustrates 7 ensembles in which two common stimulus objects are employed, and in which the range of the ensemble increases on successive rows. The 16 steps ranged from .097 sec. to 1.29 sec.

TABLE 2

SCHEMATIC PLAN OF STIMULUS ENSEMBLES USED IN EXP. III (DURATION) AND IX (INTENSITY)

Ensemble	Stimulus Objects																Derived Confusability (%)	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Exp. III	Exp. IX
1	o	o	x	x													36	33
2	o	o			x	x											19	25
3	o	o					x	x									31	
4	o	o							x	x							30	28
5	o	o									x	x					22	
6	o	o											x	x			24	
7	o	o													x	x	35	32
				4				8				12				16		
8							x	o	o	x							25	36
9						x		o	o		x						24	
10					x			o	o			x					26	
11				x				o	o				x				23	
12			x					o	o					x			21	
13		x						o	o						x		24	
14	x							o	o							x	32	
				4				8				12				16		
15													x	x	o	o	21	36
16											x	x			o	o	21	
17									x	x					o	o	23	
18							x	x							o	o	24	30
19					x	x									o	o	21	
20			x	x											o	o	17	
21	x	x													o	o	20	27
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		

Note.—In each section, two (open circles) of the four stimulus objects were present in all ensembles. The columns at the right give the derived confusability score, when available, for the common pair of stimulus objects.

TABLE 3

NUMBER OF OBSERVATIONS OF EACH STIMULUS IN EACH MATRIX AND NUMBER OF MATRICES (IN PARENTHESES) IN EACH EXPERIMENT

Experiment	Matrix Size		
	8 × 8	4 × 4	2 × 2
Contextual effects			
III (duration)		576(21)	576(3)
IX (intensity)		576(8)	576(3)
VII (intensity)[a]			
Series A	576	576(2)	576(5)
Series B	576	576(4)	576(10)
Series C	576	576(4)	576(8)
Series D	576	576(3)	576(10)
Single dimensional			
I (duration)	864	864(12)	864(14)
IV (intensity)	864	864(11)	864(14)
Multidimensional			
II (coarse)	864	864(13)	864(14)
VIII (fine)	576	576(8)	576(9)
VI (redundant)	432	864(10)	

[a] Also designed as a single-dimensional experiment.

in approximately equal logarithmic steps. The sound intensity was 80 db. spl. The observed confusability of each of the common pairs, i.e., 1, 2 in the top section; 8, 9 in the middle section; and 15, 16 in the lower section, was obtained in separate 2 × 2 tests.

Within each of the three sections, the individual ensembles of four stimuli were examined in a scrambled order. All observations associated with a given ensemble were completed before exposure of another ensemble. The number of observations of each stimulus in each 4 × 4 and each 2 × 2 confusion matrix is given in Table 3. A recognized limitation of the procedure is the possibility of inadequate counterbalancing of effects distributed in time (learning, motivation, morning vs. afternoon runs).

Results.—The derived confusability for the common stimulus pair of each ensemble is presented on the right side of Table 2. No clearly discernible trend within the large variability of the scores is observed as the range is increased.

The observed confusability (from the separate 2 × 2 tests) was 14, 25, and 24% for the common pair in the top, middle, and lower section, respectively, of Table 2.

Experiment IX: Ensemble Range

Method.—Experiment IX was identical with Exp. III except that: successive steps of intensity, spaced 1 db. apart, were employed; the duration was .32 sec.; Stimulus 1 of Table 2 corresponded to 70 db. spl; and fewer ensembles were examined.

Results.—The derived confusability scores are presented in the second right-hand column of Table 2. The observed confusability for the three common pairs was 39, 35, and 32%, respectively. As in Exp. III, the derived confusability scores do not systematically increase with an increase in ensemble range.

The RBM suggests that the derived confusability of a given stimulus pair should systematically increase as a function of the range of the stimulus ensemble. Experiments III and IX failed to observe this change, although the result must be qualified by the apparently large variability of the scores between successive conditions. Despite the large variability, the CRR describes the results reasonably well in that the derived and observed confusability are not systematically different.

Experiment VII: Ensemble Spacing

Experiment VII was designed to examine the joint effects of interstimulus spacing and ensemble range upon the derived confusability of particular pairs of stimulus objects. It also permitted an examination of the relative confusability of any pair of stimuli (and responses) as a function of the number of intermediate stimuli (and responses) between the pair.

Method.—Four interstimulus spacings were employed. They are illustrated in Table 4. Successive steps in the intensity of a 1000-cps tone, 0.32 sec. in duration, were separated by 4, 2, 1, and 0.5 db. for Series A, B, C, and D, respectively. A relative level of 0 db. represents a sound level of 70 db. spl. The

TABLE 4

SCHEMATIC ILLUSTRATION OF THE STIMULUS SPACINGS IN EXP. VII

Series	Relative Sound Level in db. 0	4	8	12	16	20	24	28
A (4 db.)	o	o	o	o	o	o	o	o
B (2 db.)	o o	o o	o o	o o				
C (1 db.)	o o o o	o o o o						
D (.5 db.)	oooooooo							

Note.—The open circles indicate the stimuli employed in each ensemble.

response label "1" was employed to designate the lowest intensity of a series; the response "8" to designate the highest intensity. Within a given series the labels employed in the 8 × 8, 4 × 4, and 2 × 2 tests were identical. However, from series to series, a given sound level was assigned a different label, e.g., a relative level of 4 db. was labeled "2" in Series A, "3" in Series B, and "5" in Series C.

Table 4 shows that the first two steps of Series A (hereafter written as A-1,2), Steps B-1,3; and Steps C-1,5 represent the same pair of sound intensities and that D-1,8 nearly represent the same pair of sound intensities. Similarly, A-1,3 and B-1,5 represent another pair of sound intensities. The number of intermediate stimuli between a given stimulus pair increases with a decrease in ensemble range. Thus, there are no intermediate stimuli (and their corresponding responses) between A-1,2; one between B-1,3; three between C-1,5; and six between D-1,8.

Each series was run in two separate blocks to permit a gross randomization of matrix size. In each block one-half of the data for the 8 × 8, 4 × 4, and 2 × 2 matrices of a given series was obtained in random order. The total number of responses to each stimulus in each matrix is given in Table 3.

Results.—The results are summarized in Table 5. The first column represents the distance in decibels between the two members of the stimulus pair under examination. The second column locates the stimuli in Table 4. The next four columns present, when available, the derived confusability, as determined from the 8 × 8 matrix (without parentheses), and the observed confusability, as determined from the 2 × 2 matrix (within parentheses). All entries in a given row represent the same stimulus pair, except those marked with an "a" in which the spacing has been shortened by 1 step-unit (i.e., 1.0 db. in C and 0.5 db. in D).

To be noted first is the lack of systematic differences as a function of the "label" assigned in the 2 × 2 tests (compare the entries within parentheses across a given line of Table 5). Stated simply, if *S* is faced with an ensemble of only two events at relative labels of 0 and 2, there is little difference if these levels represent Steps 1 and 5 in Series D, Steps 1 and 3 in Series C, or Steps 1 and 2 in Series B.

The main point of Table 5 is that the derived confusability scores decrease as the range of the initial 8 × 8

TABLE 5

OBSERVED (IN PARENTHESES) AND DERIVED CONFUSABILITY SCORES OF SELECTED STIMULUS PAIRS IN EXP. VII

Stimulus Spacing (db.)	Relative db.	Series A	B	C	D
8	0–8	5	2	0[a]	
	4–12	7	1		
4	0–4	15 (6)	8 (4)	4	2[a]
	4–8	19 (7)	10 (7)	4[a]	
	8–12	18 (8)	10 (7)		
2	0–2		19 (15)	14 (15)	13 (18)
	2–4		17	13 (14)	14[a]
	4–6		20 (20)	12 (19)	
1	0–1			21 (27)	28 (26)
	1–2			21	23 (28)
	2–3			23 (24)	23 (24)

[a] Entries of ensembles shortened by one step interval.

ensemble decreases, i.e., the entries without parentheses decrease across a given line in Table 5. Further, the systematic change in derived confusability with the range of the ensemble decreases as the spacing between the stimulus pair decreases. The change in derived confusability from Series A toward D is greatest with the 8-db. spacing and is negligible with the 1-db. spacing.

And finally, the derived confusability tends to be systematically higher than the observed confusability for the widely spaced series (Series A and B) and tends to be systematically lower for the closely spaced series (Series C and D). The greatest systematic differences, about 10% on the average, are associated with the widely spaced series (Series A).

The ensembles of Table 4 confound the total range of the ensemble with the number of intermediate stimuli (and responses). Thus, while Ensemble C has more intermediate steps than Ensemble A associated with the same stimulus pair, Ensemble C subtends a smaller range than Ensemble A. A series of 4 × 4 tests substantiated the major findings of Table 5, but, again, ensemble range was confounded with the number of intermediate stimuli.

The systematic changes of Table 5 appear to be due to a confusion among alternative responses rather than among alternative stimuli. Response confusion seems to be particularly evident in the widely spaced series (A and B) when the proximity of the response labels is close. For example, in 8 × 8 tests involving the relative sound levels of 0 and 4 db., the response labels "1" and "2" (Series A) may be confused by *S*, but the labels "1" and "5" (Series C) are less likely to be confused. In the closely spaced series, presumably confusions among stimuli overwhelm the response effect.[3]

THE CRR AND SINGLE-DIMENSIONAL ENSEMBLES

Three experiments (I, IV, VII) were designed to evaluate the CRR with stimulus objects varying along a single auditory dimension. The adequacy of the rule is tested primarily by comparing confusability scores obtained from 4 × 4 and 2 × 2 matrices with those derived from 8 × 8 matrices. The general method of the studies is the same as that previously reported.

Experiment I: Duration

Method.—The stimulus ensemble consisted of eight durations (.185-.651 sec.), equally spaced on a logarithmic scale, of a 1000-cps tone at a sound intensity of 80 db. Half of the 8 × 8 responses were obtained at the beginning, and half at the end, of the experiment. The 4 × 4 data were collected before the 2 × 2 data. The number of different 4 × 4 and 2 × 2 matrices examined and the number of observations associated with each matrix entry are given in Table 3.

Results.—Figure 1 (4 × 4) presents a comparison of the derived and observed response probabilities of the individual entries of the 4 × 4 matrices. The abscissa is the predicted proportions—the derived probabilities—from the 8 × 8 matrix; the ordinate is the obtained proportions—the observed probabilities—from the 4 × 4 tests. The CRR states that the predicted and obtained proportions should fall exactly on the 45° line of Fig. 1. Except for random variations, the rule makes excellent predictions.

A summary of the entries of Fig. 1 is presented in Table 6. The first

---

[3] An informational analysis of Exp. VII indicated that the amount of information achieved with each stimulus series (relative to the maximum possible transmitted information) is independent of both the spacing within each stimulus series and the number of stimulus objects. The amount of transmitted information seems to be determined only by the range subtended by the stimulus series.

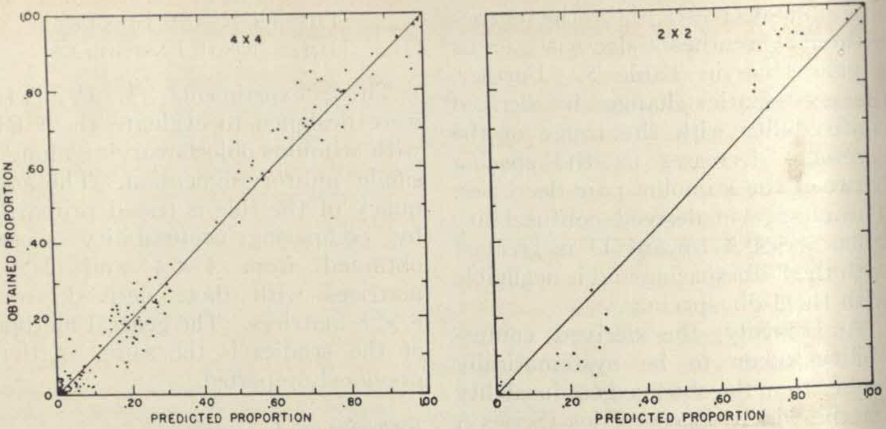


FIG. 1. Predicted and obtained response proportions of the individual 4 × 4 and 2 × 2 entries in Exp. I.

column gives the size of the matrix: 8 × 8, 4 × 4, or 2 × 2. The second column is the mean proportion of obtained correct responses. The third column is the mean of the differences (sign ignored) between the predicted and obtained proportions. The last column gives the mean systematic (sign not ignored) difference between the obtained and predicted proportions of the correct responses. A positive entry indicates that the obtained proportion was greater than the predicted proportion. In addition the obtained proportion correct scores of the 4 × 4 and 2 × 2 matrices were rank ordered, and then, on the basis of the ranks, the 4 × 4 and 2 × 2 matrices were divided equally into high and low proportion groups (first and second half, Table 6). The various difference measures were then calculated for the separate groups.

The entries of Table 6 (Exp. I) confirm the excellent CRR predictions of the 4 × 4 matrices shown in Fig. 1. The mean of absolute differences between the predicted and obtained scores is .028, and 4.2% (not given in Table 6) of the differences, on the average, are larger than .10, the arbitrary criterion used by Clarke (1957). The mean of the systematic differences is +.014, indicating that the obtained proportions are slightly larger than the predicted proportions, i.e., the derived confusability is greater than the observed confusability.[4] A comparison of the difference measures of the two halves of the 4 × 4 matrices indicates that somewhat better predictions are obtained with the matrices of the first half.

Table 6 and Fig. 1 also show that the CRR provides questionable predictions of the 2 × 2 matrices. The mean of the absolute differences is .082 and the average percentage of differences greater than .10 is 28.6. The mean systematic difference (+.052) suggests that the derived confusability is again greater than the observed confusability. This finding is graphically illustrated by

---

[4] In calculating the systematic differences for a given matrix, it was more convenient to use the proportion of correct responses than the proportion of incorrect responses. Confusability scores can be obtained by subtracting the proportion of correct responses from 1.0.

TABLE 6

SUMMARY OF RESULTS OF SINGLE-DIMENSIONAL AND MULTIDIMENSIONAL EXPERIMENTS

Matrix Size	Mean Prop. Correct	Mean |Diff.|	Mean Diff. of Diagonal Entries
Experiment I: Duration			
8 × 8	.568		
4 × 4	.746	.028	+.014
1st half	.784	.022	−.003
2nd half	.707	.034	+.032
2 × 2	.933	.082	+.052
1st half	.992	.057	.000
2nd half	.874	.107	+.105
Experiment IV: Intensity			
8 × 8	.713		
4 × 4	.831	.025	+.013
1st half	.860	.023	+.001
2nd half	.802	.027	+.025
2 × 2	.971	.043	+.042
1st half	.998	.003	.000
2nd half	.944	.083	+.084
Experiment II: Coarse Spacing			
8 × 8	.794		
4 × 4	.908	.020	+.020
2 × 2	.986	.023	+.018
Experiment VIII: Fine Spacing			
8 × 8	.485		
4 × 4	.666	.028	+.002
2 × 2	.846	.038	−.004
Experiment VI: Redundant Objects			
8 × 8	.834		
4 × 4	.947	.018	+.028

the clustering of the entries on one side of the diagonal in Fig. 1. The difference measures are especially large in the matrices with the lower proportion of correct responses (Table 6). The accuracy of prediction thus seems to be strongly dependent upon the obtained proportion of correct responses. The largest differences are always found in matrices with adjacent stimulus (and response) objects, e.g., 1,2; 5,6.

Experiment IV: Intensity

Method.—Experiment IV was identical with Exp. I except that an ensemble of eight intensities, spaced 3 db. apart, was examined. The lowest sound level was 70 db. spl. The frequency of the tone was 1000 cps and the duration was .32 sec. Table 3 gives the number of observations of each stimulus in the 8 × 8, 4 × 4, and 2 × 2 matrices.

Results.—Figure 2 shows a close correspondence between the obtained and predicted response probabilities for the 4 × 4 matrices. The mean of the absolute differences is .025 (Table 6) and the mean percentage of differences greater than .10 is 2.8. An analysis of the systematic differences between the predicted and obtained probabilities for the correct responses produced a mean difference of +.013. Thus, all the difference measures support the accuracy of the CRR predictions shown in Fig. 2. A comparison of the difference scores for the first and second halves of the 4 × 4 matrices indicates that the CRR predicts equally well the entries of matrices with high and lower proportion correct scores. This confirms a similar result obtained in Exp. I (Duration).

The predictions of the CRR are much poorer for the 2 × 2 data (Fig. 2 and Table 6). The mean absolute difference is .043; 21.5% of the differences exceed .10, on the average; and the mean of the systematic differences for the correct responses is +.042. An analysis of the two halves of the matrices (Table 6) indicates that the difference measures for the matrices with high proportion correct scores approach zero while those of the second half are larger. As in Exp. I, the largest differences are associated with matrices made up

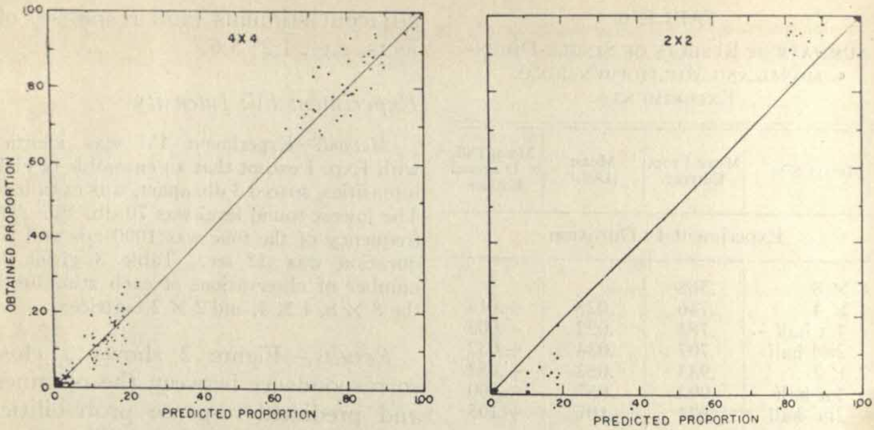


FIG. 2. Predicted and obtained response proportions of the individual 4 × 4 and 2 × 2 entries in Exp. IV.

of adjacent stimulus (and response) objects.

Experiment VII: Intensity Spacing

Method.—The method of Exp. VII was presented in the section concerned with contextual effects. Only the results relevant to the CRR will be presented here.

Results.—Table 7 summarizes the results for the four series of Exp. VII. Its organization is identical with that of Table 6. The mean difference measures for the 4 × 4 matrices demonstrate reasonably good CRR prediction. However, the mean of the diagonal differences (systematic effects) indicates that the rule underestimates the mean predicted proportion of correct responses for the widely spaced series (A and B) and overestimates those for the closely spaced series (C and D).

Inspection of the results of the 2 × 2 tests (Table 7) suggests that the CRR predictions are again reasonably good, with the possible exception of Series A (the widest ensemble.) The difference measures of Series A are all larger than those of the other series. Again, it should be noted that the largest differences are observed with matrices containing adjacent stimulus (and response) objects.

TABLE 7

SUMMARY OF RESULTS OF EXP. VII (INTENSITY SPACING)

Series	Mean Prop. Correct	Mean |Diff.|	Mean Diff. of Diagonal Entries
8 × 8			
A (4 db.)	.746		
B (2 db.)	.597		
C (1 db.)	.527		
D (0.5 db.)	.375		
4 × 4			
A	.852	.034	+.053
B	.748	.025	+.006
C	.629	.041	−.068
D	.529	.042	−.006
2 × 2			
A	.939	.088	+.084
B	.904	.034	+.022
C	.803	.051	−.030
D	.756	.054	−.029

Discussion

The serious failures of the CRR are associated with adjacent steps where the stimulus (and response) spacing is wide. In these instances the predicted proportions are seriously underestimated relative to the obtained proportions. The result is not unreasonable when examined in terms of stimulus and response confusions. A pair of stimulus objects, separated by a large stimulus difference, may be easily discriminated in a 2 × 2 test. However, within an 8 × 8 matrix, the same stimulus objects and their associated responses may be thoroughly confused. Presumably, the confusion is primarily between responses rather than between stimuli. The confusion among responses is most clearly evident in Series A of Exp. VII. Since the stimuli of Series A are more widely spaced than those of the other series, it seems inappropriate to assign the relatively large differences of Series A (especially those of the 2 × 2 data) to stimulus confusion. The other series of Exp. VII do not show such effects presumably because confusions among the stimuli of these series overwhelm any response effect. Furthermore, the overestimation of the predicted proportions in the closely spaced series (C and D) of Exp. VII suggests that the effect of stimulus confusions not only overwhelms but tends to oppose the response effect.

The importance of interstimulus spacing is shown more clearly in Fig. 3 which provides a further analysis of Exp. I. Derived confusability scores from the 8 × 8 matrix and from appropriate 4 × 4 matrices are plotted along with observed confusability scores from 2 × 2 matrices. The abscissa is the spacing between the members of selected stimulus pairs (and their responses) within a matrix; the ordinate is the mean proportion of incorrect responses (confusability score) for all pairs with the same interstimulus (and response) spacing, e.g., the Pairs 1 and 4 and 2 and 5 have the same spacing. The main feature of Fig. 3 is that the derived scores (from 8 × 8 and 4 × 4 matrices) are systematically higher than the observed scores (from 2 × 2 matrices) only when the interstimulus (and response) spacing is small, i.e., when the stimulus (and response) objects are adjacent or nearly adjacent. A similar analysis of Exp. IV also showed that the largest differences between observed and derived confusability occur in matrices with adjacent items. That the differences are due to confusion among responses rather than among stimuli is suggested by the widely spaced ensembles used in the experiments.

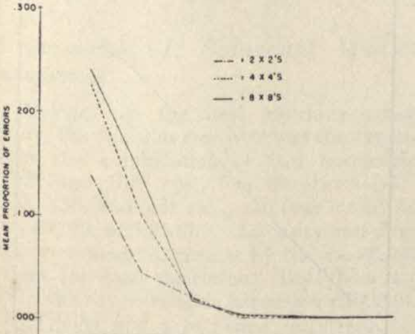


FIG. 3. Confusability scores in Exp. I as a function of interstimulus distance.

Finally, the accuracy of the CRR predictions must necessarily be a function of the relative performance level of *S*s. Excellent predictions are always obtained when the proportion of correct responses approaches 1.00. A comparison of the results given in the first half and the second half of Table 6 (Exp. I and IV) shows that in every case the smallest differences are associated with matrices in which a relatively high proportion of correct responses was obtained.

THE CRR AND MULTIDIMENSIONAL ENSEMBLES

The final experiments (II, VIII, VI) were designed to test the CRR with stimulus objects which vary along three stimulus dimensions. These ensembles represent test conditions between the single-dimensional displays described earlier in the present report and complex displays such as

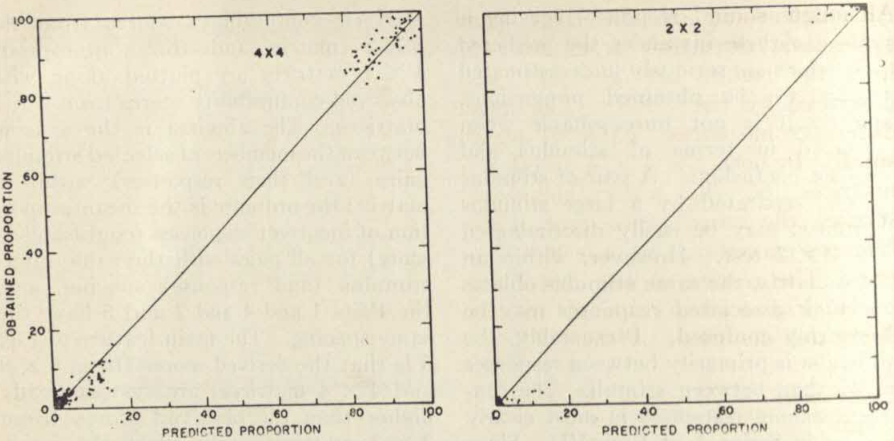


FIG. 4. Predicted and obtained response proportions of the individual 4 × 4 and 2 × 2 entries in Exp. II.

speech. The experiments examine CRR predictions as a function of interstimulus spacing and stimulus redundancy. The general method of the studies is the same as that in the single-dimensional experiments.

Experiment II: Coarse Multidimensional

Method.—An ensemble of eight stimulus objects was constructed from the combination of two tonal frequencies (1000 or 1025 cps), two durations (.255 or .367 sec.), and two sound levels (80 or 84 db.). For example, Stimulus 1 consisted of a tone 1000 cps in frequency, .255 sec. in duration, and 80 db. in intensity. Since the two values of each dimension were rather widely spaced, the ensemble is described as coarse. In all of the present multidimensional experiments, *S*s' push buttons were labeled with respect to all three dimensions. For example, the button for Stimulus 1 was labeled: low, short, soft. Table 3 summarizes the number of responses obtained in the various matrices.

Results.—Figure 4 presents a comparison of the predicted and obtained proportions of the 4 × 4 matrices.

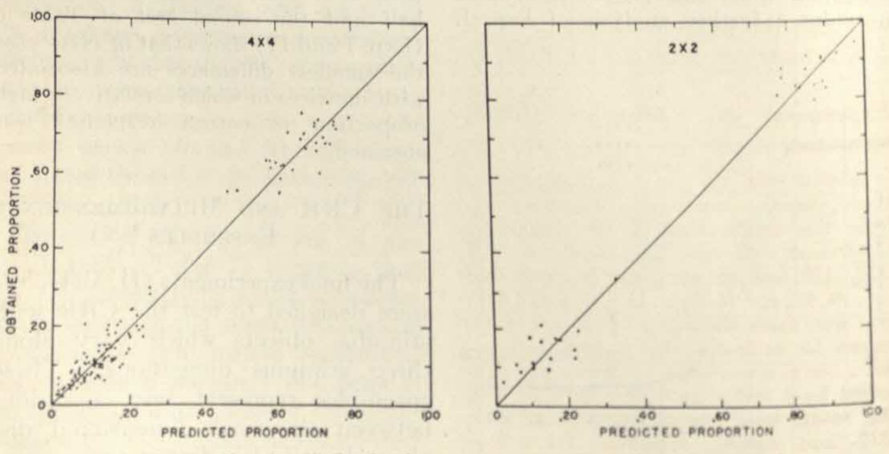


FIG. 5. Predicted and obtained response proportions of the individual 4 × 4 and 2 × 2 entries in Exp. VIII.

Although some of the data show systematic deviation from the 45° line, the results are generally well described by the CRR. The mean difference measures (Table 6) are all small. In both the 4 × 4 and 2 × 2 matrices of the multidimensional studies, the mean percentage of differences exceeding .10 was less than two. The results of the 2 × 2 matrices (Fig. 4 and Table 6) are also fairly well described by the CRR.

Experiment VIII: Fine Multidimensional

Method.—The eight stimulus objects varied with respect to: a frequency of 1000 or 1006 cps, a duration of .320 or .367 sec., and an intensity of 80 or 81 db. Experiments II and VIII were otherwise identical except that data collection followed the block system (method of counterbalancing) of Exp. VII. The number of presentations of each stimulus in each ensemble is given in Table 3.

Results.—The close correspondence between the obtained and predicted proportions of the 4 × 4 matrices (Fig. 5) is quantitatively confirmed by the low values of the difference measures presented in Table 6. Comparable high agreement between the obtained and predicted proportions was also the case for the data of the 2 × 2 matrices (Fig. 5 and Table 6).

Experiment VI: Redundant Multidimensional

Method.—In the final multidimensional study, the stimulus ensemble was constructed from the combination of two frequencies (1000 and 1025 cps), four durations (.237, .320, .450, and .651 sec.), and four intensities (86, 89, 92, and 95 db.). Intensity and duration were made redundant by the use of four values for each dimension. But there was not a one to one relation between a particular sound level and a particular duration, e.g., the 86-db. level was paired with both the .237- and .450-sec. durations. Data were collected only for the 8 × 8 matrix and 10 4 × 4 matrices (Table 3).

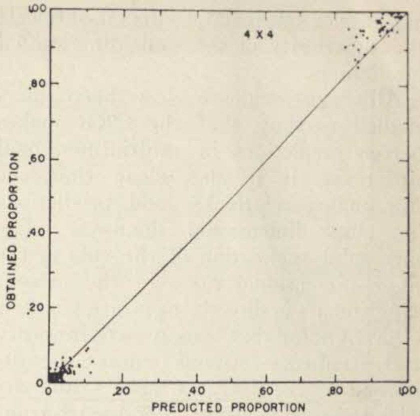


FIG. 6. Predicted and obtained response proportions of the individual 4 × 4 entries in Exp. VI.

Results.—The results are shown in Fig. 6 and Table 6. Although there is good agreement between the obtained and predicted proportions, there is a definite tendency for the points to deviate systematically from the 45° line. The extent of the deviations is more precisely indicated by the relatively large mean of diagonal differences (+.028).

Discussion

In all of the multidimensional experiments the CRR makes good or excellent predictions of the response probabilities of the various 4 × 4 and 2 × 2 matrices. Furthermore, the rule seems to hold under a variety of conditions in the multidimensional situation, e.g., coarseness and fineness of the display, the presence or absence of stimulus redundancy.

Perhaps a more important finding, however, is that the CRR predicts the response probabilities of the multidimensional matrices somewhat better than those of the single-dimensional matrices. For example, in eight of the single dimensional and in none of the multidimensional tests, the mean difference scores exceed .04. A similar analysis of the mean difference of diagonal entries (systematic effects) supports the superiority of the multidimensional predictions.

Although evidence has been marshalled to show that the CRR makes better predictions in multidimensional situations, it is also clear that the rule makes relatively good predictions for single-dimensional displays. The successful application of the rule in the single-dimensional cases of the present experiments is directly opposite Clarke's (1959) finding that responses to intensity and frequency stimuli cannot be described by the CRR. Clarke's difficulty with the rule, however, was due to strong contextual effects present in his data.[5]

SUMMARY

Eight experiments were designed to evaluate a model—the constant-ratio rule—of sensory organization. The model was tested by comparing the obtained probabilities of 4 × 4 and 2 × 2 confusion matrices with the probabilities predicted from 8 × 8 matrices. The auditory dimensions of frequency, intensity, and duration were used singly and in combination to form single and multidimensional stimulus ensembles consisting of two, four, or eight tones. Six *S*s attempted to identify, one at a time, the stimulus objects of the ensembles.

The results indicate: (*a*) In general, the CRR makes excellent predictions of the response probabilities of the single and multidimensional matrices. The notable exceptions occur when predictions for 2 × 2 matrices are made from adjacent objects in widely spaced ensembles. (*b*) The CRR predicts the probabilities of the multidimensional matrices somewhat better than those of the single-dimensional ensembles.

REFERENCES

ANDERSON, C. D. The constant-ratio rule as a predictor of confusions among visual stimuli of brief exposure duration. *USAF Cambridge Res. Cent. tech. Note,* 1959, No. TN 58-60. (ASTIA No. AD-160706).

CLARKE, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. Acoust. Soc. Amer.*, 1957, **29,** 715–720.

CLARKE, F. R. Proportion of correct responses as a function of the number of stimulus-response alternatives. *J. Acoust. Soc. Amer.*, 1959, **31,** 835. (Abstract)

CLARKE, F. R., & ANDERSON, C. D. Further test of the constant-ratio rule in speech communication. *J. Acoust. Soc. Amer.*, 1957, **12,** 1318–1320.

McGILL, W. J. The slope of the loudness function: A puzzle. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling.* New York: Wiley, 1960. Pp. 67–81.

POLLACK, I., & DECKER, L. Consonant confusions and the constant ratio rule. *Lang. Speech,* 1960, **3,** 1–6.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

(Received November 12, 1960)

---

[5] F. R. Clarke, personal communication, 1960.

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 143–147

REACTION TIME TO STIMULI MASKED BY METACONTRAST[1]

ELIZABETH FEHRER AND DAVID RAAB

Brooklyn College

The experiment to be described below is concerned with the relation between (*a*) phenomenal characteristics and (*b*) an objective behavioral measure of metacontrast suppression. If a brief light flash is followed, after a suitable interval, by stimulation of adjacent retinal areas, its apparent brightness is reduced. The extent of such retroactive masking has been shown to be a U shaped function of stimulus onset asynchrony, with the maximum suppression occurring at asynchronies of 70 to 100 msec. (e.g., Alpern, 1953). At such asynchronies, the first stimulus may be phenomenally absent (Kolers & Rosner, 1960; Toch, 1956; Werner, 1935). It is not, however, without effect on perception since under these conditions there is an awareness of movement involving the second stimulus. (Such asynchronous stimulation of adjacent retinal areas is, of course, the condition for producing apparent movement.)

The problem of the present experiment was to determine whether the phenomenal masking of the first stimulus is associated with a decrease in its capacity to elicit a simple overt response. Specifically, we hoped to determine whether reaction time (RT) is correlated with the apparent brightness of a flash of light subjected to various degrees of metacontrast masking. The question becomes of especial interest at those stimulus onset asynchronies at which the first flash is phenomenally absent and only the masking stimuli (flashed second) are perceived.

The present problem is related to that raised in a previous study (Raab, Fehrer, & Hershenson, 1961) which dealt with the relation between RT and the apparent brightness of light flashes as contingent on their duration. It was shown that RT depended on stimulus characteristics rather than on the phenomenal appearance of the stimulus.

RT is a convenient measure of retroactive masking since the masked stimulus is the one presented first. If the RT chronoscope is started when the first stimulus is presented, it is possible to compare the RT to this stimulus when it is presented alone with the RT to it when it is followed by masking stimulation. Any lengthening of RT can then be attributed to the masking procedure.

METHOD

Apparatus

The stimulus lights and *S* were located in one room; *E*'s control panel and the RT chronoscope were in another room.

The stimulus lamps were placed behind a 6 × 2 in. rectangular box divided by two partitions into three 2 × 2 in. square cells arranged horizontally (see Fig. 1). The top, bottom, sides, and septa of the box were made of sheet aluminum, $\frac{1}{32}$ in. thick. The front of the array (as seen by *S*) was covered with a sheet of diffusing tissue pressed against the cell walls by a 18.5 × 8.5 in. pane of $\frac{1}{8}$-in. polished plate glass mounted in a 2-in. wooden frame. The entire assembly was placed at eye level and was viewed binocularly by *S* from a distance of 15 ft. The horizontal visual angle of the array was 2°. For foveal stimulation, two pin-point red lights were placed at the locations marked F (Fig. 1),

---

[1] This research was supported by Grant G-6456 from the National Science Foundation.

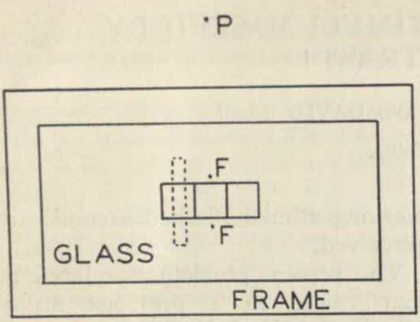


FIG. 1. The stimulus array as seen by *S*. (The three light cells are covered by a sheet of diffusing tissue. The position of the fluorescent tube in each cell is indicated by the dashed outline in the left-hand cell. F and P are fixation points. For dimensions, see text.)

and *S* was instructed to fixate between them. Autokinetic movement of these fixation lights in the otherwise dark test room was eliminated by dimly illuminating the frame surrounding the plate glass. This illumination was too faint to be measured. The frame did not become visible to *S* until after 5 min. of dark adaptation.

For peripheral stimulation, a single fixation light, P (Fig. 1), was placed 8.75 in. (i.e., 2° 50′) above the top center of the central light cell.

A cold cathode fluorescent lamp was placed at the rear of each of the three light cells with phosphored portion 7¼ in. behind the diffusing tissue viewed by *S*. The lamps are of the mercury vapor type and are coated with a calcium halo phosphor. Luminance, measured at the diffusing tissue, was approximately 18 ft-L.

The lamps were pulsed by a "hard tube" high-voltage circuit. The vacuum tubes in series with each lamp were driven by a Tektronix 161 pulse generator. For all the work reported here, the lamp in the center cell was pulsed by one circuit and the two flanking cells were flashed together by an independent circuit. Regulated power supplies were employed throughout. When flashed, the lamps provide reasonably rectangular pulses of light. The chief departure occurs at offset and results from phosphor persistence.

Only two durations of flash, 5 and 50 msec., were used in this experiment. Since the light output from a lamp increases slightly with durations beyond 5 msec., the luminances of the two pulses at offset were not equal. The 50-msec. pulse reached a level 0.076 log units above that of the 5-msec. flash.

Timing of the stimulus pulses was accomplished by means of a set of Tektronix pulse and wave form generators. All critical timing voltages were regulated and the times themselves were monitored on a Westport WE-210 counter.

RT trials were presented once every 15 sec. Each trial began with a 1-sec. warning tone. The stimulus pattern was presented (and the counter started) 2.9, 3.2, or 3.5 sec. after the onset of the tone. The three foreperiods were switch selected and were presented in random order. The counter was stopped when *S* depressed a normally closed telegraph key.

RTs to the center light alone and to the center-plus-flanks could be read directly from the counter. RTs to the flanks alone were computed by subtracting the stimulus asynchrony from the counter readings.

Procedure

Preliminary exploration indicated that good metacontrast suppression could be obtained with various durations of the center and flanking lights. It was arbitrarily decided to limit the present investigation to the following combinations: (*a*) center, 50 msec.; flanks, 50 msec. (50-50-50); (*b*) center, 5 msec.; flanks, 50 msec. (50-5-50). Each combination was studied in foveal and in peripheral vision.

Preliminary.—Systematic variation of stimulus onset asynchrony (Δt) yielded the following perceptual correlates of asynchrony for the 50-50-50 sequence with foveal vision. The first perceptible darkening of the center light (as compared with the flanks) appears at a Δt of 4 to 6 msec. The darkening progresses with increases in Δt to about 35 msec. At this asynchrony, the center field is no longer homogeneous in color nor phenomenally a square, but appears as a faint, shapeless flash of light in the center of an extremely dark field between the flanks. Maximum suppression occurs with a Δt of approximately 75 msec. With increases in Δt beyond about 90 to 100 msec., depending on *S*, the center flash brightens again. At approximately 120 msec., it appears as a square preceding the flanks and of the same brightness as the flanks. Its apparent duration, however, is less than that of the flanks. With a Δt of about 150 msec., both brightness and duration of the center seem the same as those of the flanks.

With the 50-5-50 lights in foveal vision, the center square does not appear homogeneous even with a Δt of zero. Instead, only its center appears light. Maximum darkening and reappearance occur at about the same Δt's as for the 50-50-50 lights.

In peripheral vision, the phenomenal correlates of asynchrony differed from those described above for foveal vision only in that the apparent darkening of the center light was more pronounced than with foveal vision. With Δt's from 50 to 100 msec., the entire center square area appeared completely dark.

On the basis of the above findings, the following asynchronies were selected for systematic study: 0, 10, 25, 50, and 75 msec. It was assumed that, if RT parallels apparent brightness of the center light, RT, measured from the onset of the center light, should increase with Δt, reaching a maximum at a Δt of 75 msec., and that here, especially in peripheral vision, it might well be the same as RT to the flanks.

Foveal vision.—The authors and a graduate student served as *S*s. All had had extensive previous experience with RT. Each *S* served in 50 sessions, 25 with the 50-50-50 durations and 25 with the 50-5-50. Of the 25 sessions at a given duration of the center light, 5 were assigned to each of the 5 stimulus onset asynchronies.

Thirty-six RTs were recorded at each session, 12 to the center light alone, 12 to the flanks alone, and 12 to the combination. Random presentation prevented *S* from knowing which stimulus pattern would be presented at a given trial. Foreperiod duration was also randomized.

The longest and the shortest RT of each 12 were eliminated before the three means for each session were computed. Since there were five sessions for each asynchrony, the means for each *S* are based on 50 RTs.

Two sessions were run on each experimental day, one with the 50-50-50 and one with the 50-5-50 durations. Sessions at the various Δt's followed each other in random order.

After 5 min. of dark adaptation, three practice trials were given. Then the recorded trials were run. A 1-min. rest period separated each 18 trials.

Peripheral vision.—For the 50-50-50 durations, two graduate students and one undergraduate major served as *S*s. For the 50-5-50, one author and one undergraduate major were *S*s. Each served in 20 sessions, 5 each at the following four Δt's: 0, 10, 25, and 75 msec. The procedure was otherwise identical with that employed for foveal vision.

TABLE 1

MEAN REACTION TIMES WITH FOVEAL STIMULATION

Δt in Msec.	50-50-50			50-5-50		
	Center	Flanks	All	Center	Flanks	All
0	165.9	163.3	161.1	167.3	162.3	162.6
10	164.6	162.3	161.9	166.2	162.6	161.3
25	167.1	163.4	164.9	169.6	166.6	167.1
50	167.2	166.6	167.5	166.5	162.5	166.5
75	162.4	160.3	164.5	165.4	161.9	165.1
Mean	165.4	163.2		167.0	163.2	

RESULTS AND DISCUSSION

Since all *S*s showed the same trends, only the group means are presented in Tables 1 and 2. The individual *SD*s were of the order of 12 msec. and did not differ with stimulus conditions. The data in both tables show that RT to the center light was never *increased* by the addition of the masking stimuli. A series of *t* tests of the difference between each individual center and center-plus-flanks means showed that in no case was the RT to center-plus-flanks reliably longer. It follows, therefore, that decreases in apparent brightness produced by the metacontrast procedure are not reflected in RT increases. This is true even for peripheral vision where, with a Δt of 75 msec., the center flash is phenomenally absent. Under all our conditions, RT seemed to be determined only by the *stimulus*

TABLE 2

MEAN REACTION TIMES WITH PERIPHERAL STIMULATION

Δt in Msec.	50-50-50			50-5-50		
	Center	Flanks	All	Center	Flanks	All
0	159.6	155.2	153.4	167.6	160.2	160.9
10	162.3	155.6	158.9	164.4	160.0	163.5
25	163.2	160.5	164.1	164.8	159.6	164.3
75	161.8	158.1	161.5	165.7	160.0	165.0
Mean	161.8	157.3		165.6	159.9	

characteristics of the light flashed first.

The fact that the stimulus does evoke a normal RT even when completely "masked" is not, perhaps, entirely unexpected. Since the metacontrast procedure is also the procedure for producing phi movement, it is obvious that the light flashed first does affect the total perception. Even though the center light appears "blanked," its prior occurrence is a necessary condition for the perception of movement involving the flanks, and such movement was apparent to all our *S*s. Our data have shown clearly that RT to a light is the same when (*a*) it is presented alone and perceived as a bright flash, and when (*b*) it is followed by later stimulation of adjacent retinal areas and is "perceived" only through the apparent movement of these lights.

Our results do not depend on knowledge of stimulus conditions. RTs of 6 naive *S*s were measured to the center light alone and to the center followed after 75 msec. by the flanks. The 50-50-50 lights and peripheral fixation were used. Although the RTs were longer than those appearing in our tables (*S*s were unpracticed), they were the same under both stimulus conditions. When these *S*s were asked to describe what they saw, all reported seeing either a single square of light or two squares separated by a dark area about the size of one of the squares. Movement of the two squares was never spontaneously reported until at least 20 RT trials had been run.

The *S*s were then tested to see whether they could learn to distinguish phenomenally between the appearance of the flanks alone and the center-plus-flanks. All but 1 learned to do so but not until 10 to 20 trials (with correction) had been run.

In other words, these *S*s' RTs were from the beginning a reliable index of the presence of the "masked" center light whereas their phenomenal reports became adequate only after some specific training.

The data provide some information concerning the effect of stimulus duration on RT. The center stimulus was presented alone on one-third of the trials in each session. Since both 5-and 50-msec. flashes were employed, RTs to these can be compared directly. With foveal vision (Table 1), the mean RT to the 50-msec. flash was 165.4 msec.; the mean RT of the same *S*s to the 5-msec. pulse was 167.0 msec. The absence of any marked difference in RT is in line with the results of a previous study (Raab et al., 1961) in which RT was found to be independent of stimulus durations between 10 and 500 msec. A similar comparison cannot be made with the present peripheral data since the 5- and 50-msec. flashes were presented to different *S*s.

That stimulus size affects RT (e.g., Froeberg, 1907) is shown by our data. Reactions to the single 50-msec. center flash were slower than to the two 50-msec. flanks. Of the 15 possible individual comparisons between center and flank means for foveal vision (3 *S*s × 5 asynchronies), 13 show faster RT to the flanks than to the center ($P = .01$). Of the 12 comparisons (3 *S*s × 4 asynchronies) in peripheral vision, 11 show faster RT to the flanks ($P < .01$). For foveal vision, the overall mean of 165.4 msec. for the center light differs significantly ($P < .01$) from the mean of 163.2 for the flanks (Table 1, last row). The comparable difference for the peripheral data is also significant.

When the center and flanks are pulsed simultaneously ($\Delta t = 0$), the stimulus area is, in effect, three times that of the center alone. RT to this largest stimulus was in all cases faster than RT to the flanks.

It is interesting to note that RT to the center-plus-flanks is also faster than RT to the center alone when the flanks are delayed by 10 msec. Of the 11 differences between individual means at this asynchrony (6 for foveal and 5 for peripheral vision), 10 show this effect ($P = .01$). There is some indication that this effect persists up to the 25-msec. Δt for foveal vision, since in the six possible comparisons under this condition, RT to the combined stimulation was faster in all cases than RT to the center alone. The finding of "facilitation" at these two asynchronies is the opposite of what would have been found if phenomenal brightness governs RT since, at these asynchronies, the center flash is appreciably darkened. The finding is similar to an intersensory facilitation of RT recently reported by Hershenson (1960).

SUMMARY

This study was designed to determine whether metacontrast suppression of a light flash affects RT to the flash. Masking of a square target was achieved by subsequent flashing of two adjacent squares. Stimulus onset asynchronies were studied over a range from 0 to 75 msec. Phenomenal suppression of the first flash varied from none (0 asynchrony) to maximum (75 msec. asynchrony).

With both foveal and peripheral vision, the results showed no effect of phenomenal brightness on RT. RT to the target, flashed alone, was never faster than RT to the target followed by masks. It was therefore the physical dimensions of the stimulus rather than its phenomenal characteristics that determined RT.

REFERENCES

ALPERN, M. Metacontrast. *J. Opt. Soc. Amer.*, 1953, **43**, 648–657.

FROEBERG, S. The relation between the magnitude of stimulus and the time of reaction. *Arch. Psychol., NY*, 1907, **16**, 1–38.

HERSHENSON, M. Reaction time as a measure of intersensory facilitation. Unpublished master's thesis, Brooklyn College, 1960.

KOLERS, P. A., & ROSNER, B. S. On visual masking (metacontrast): Dichoptic observation. *Amer. J. Psychol.*, 1960, **73**, 2–21.

RAAB, D., FEHRER, E., & HERSHENSON, M. Visual reaction time and the Broca-Sulzer phenomenon. *J. exp. Psychol.*, 1961, **61**, 193–199.

TOCH, H. The perceptual elaboration of stroboscopic presentations. *Amer. J. Psychol.*, 1956, **69**, 345–358.

WERNER, H. Studies on contour: I. Qualitative analysis. *Amer. J. Psychol.*, 1935, **47**, 40–64.

(Received November 15, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 148–154

THE EFFECT OF POSTURAL FACTORS ON THE DISTRIBUTION OF TACTUAL SENSITIVITY AND THE ORGANIZATION OF TACTUAL-KINAESTHETIC SPACE[1]

JOSEPH H. McFARLAND, HEINZ WERNER, AND SEYMOUR WAPNER

Clark University

The distribution of tactual sensitivity has been widely held to correspond with the spatial distribution of touch receptors on the body: where there is a high density of receptors there is high sensitivity. There are three lines of evidence however, which question the plausibility of this strictly peripheral explanation of the distribution of tactual sensitivity first proposed by Weber (1846). First, on anatomical grounds it is questionable as to whether there is a strict one-to-one correspondence between the spatial distribution of touch receptors and the distribution of tactual sensitivity (Jenkins, 1951). Second, it is questionable as to whether high density of receptors is the sole condition for high sensitivity, since sensitivity depends on judgmental factors (Boring, 1942, pp. 480–483; Gibson, 1953). Third, it is questionable as to whether tactual sensitivity can be explained solely in terms of the tactual afferent system, since studies with cerebellar patients suggest that sensitivity also depends on the postural status of the organism (Halpern & Kugelmas, 1957).

The first aim of this paper is to consider further the question of whether tactual sensitivity can be explained solely in terms of the tactual afferent system by systematically investigating the effect of postural changes on sensitivity in normal adults.

The second aim of this paper is to generalize from previous findings in the visual modality to the tactual modality. Previously McFarland (1957) and McFarland, Werner, and Wapner (1960) have shown that changes in the pattern of proprioceptive stimulation from the oculomotor system lead to a redistribution of visual sensitivity. With increased convergence of the right eye during binocular viewing, there is an increase in acuity to the right of the focal point; conversely, with increased convergence of the left eye during binocular viewing, there is an increase in acuity to the left of the focal point. Concomitant with these changes in acuity there is a reorganization of visual space: with increased convergence of the right eye during binocular viewing, there is a shift of the apparent median plane (the straight-ahead) to the right of the objective median plane; conversely, with increased convergence of the left eye during binocular viewing, there is a shift of the apparent median plane to the left of the objective median plane (Bruell, 1953). The question to be investigated is whether there is a concomitant redistribution in tactual sensitivity and reorganization

---

[1] This investigation was supported by a USPHS grant, M-348, from the National Institute of Mental Health, United States Public Health Service. The authors wish to thank David Calverley who served as Research Assistant for some of the experiments.

in tactual-kinaesthetic space depending on postural changes.

Two experiments are reported: the first deals with the effect of postural changes on the distribution of tactual sensitivity and the second deals with the effect of postural changes on the organization of tactual-kinaesthetic space.

EXPERIMENT I

Generalizing from the earlier findings concerning a redistribution in visual sensitivity, the following hypothesis was formulated: tilt left increases sensitivity of the right palm compared to the left palm; tilt right increases sensitivity of the left palm compared to the right palm. Three sets of postural conditions—head tilt, supported body tilt, and unsupported body tilt—were used, rather than only one set, in order to increase the breadth of the findings with respect to the predicted relationship between postural changes and tactual sensitivity changes.

Method.—For each *S*, a tactual two-point threshold was obtained for left and right palms. Independent groups of *S*s were used for each set of postural conditions. Each set of postural conditions consisted of three positions: tilt 20° left (L), erect (E), and tilt 20° right (R). These variations within a set were arranged in a 3 × 3 replicated latin square. For each sequence, half the *S*s were tested first for left palm and then for right palm; the remaining half were tested first for right palm and then for left palm. Thus for each *S* a tactual two-point threshold was obtained for left and right palms under a set of postural conditions.

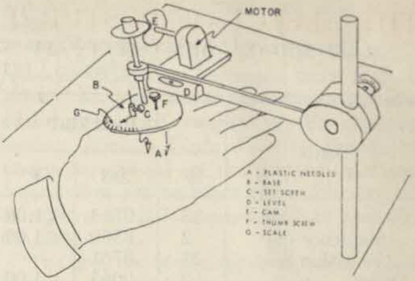


FIG. 1. Schema of aesthesiometer used to determine two-point threshold.

The important features of two aesthesiometers (Fig. 1) constructed for this experiment are as follows: (*a*) Two plastic needles (A) are mounted in sleeves set at right angles to a base (B). Set screws in the sleeves (C) permit independent variation of the height of the needles. Thus, the needles can be adjusted to compensate for irregularities of the palm's surface so that the needles strike the palm at the same time and with equal pressure. Total weight of the needles and base is 22 gm. (*b*) The needles may be oriented by means of a level (D) so that they fall perpendicularly. (*c*) By means of a motor driven cam (E), the needles drop onto the palm, remain there for 2 sec., then are raised and after 23 sec. drop again. (*d*) Distance between the needles can be varied by means of a thumb screw and gear arrangement in the base (F). Distance between the needles in millimeter units can be read directly from a scale on the base (G).

Each S extended his forearms at right angles to his trunk resting the dorsal side of the hands on individually fashioned Plasticine moulds. The *S*'s arms and hands were kept in this position throughout the experimental session. With *S*'s eyes closed, a two-point threshold was determined by the method of limits using one ascending and one descending trial. In the ascending series the points were started 2 mm. apart. In the descending series the points were started 20 mm. apart. Successive trials in each series differed by 2-mm. steps. The *S* was restricted to two categories of judgment: "one-point" or "two-points"; he was instructed to use the category "two-points" only when he clearly experienced two points. There was a 23-sec. interval between each trial in a series. After a threshold was determined for one palm, this procedure was repeated *immediately* for the other palm.

In the analysis, the log of the difference between thresholds for left minus right palms plus a constant was utilized.

In Cond. 1 (Head Tilt) *S*, with body erect, assumed three head positions: L, E, and R. A scale behind *S* assured that he maintained the required head position. Hands and arms were kept in the same position relative to *S*'s body for all positions. Twenty-four *S*s were tested. In Cond. 2 (Supported Body Tilt) *S* stood in an open box (padded with foam rubber to keep him straight) which could be

TABLE 1

EXPERIMENT I: ANALYSES OF VARIANCE FOR LOG THRESHOLD DIFFERENCE SCORES BETWEEN LEFT AND RIGHT PALMS

Source	1. Head Tilt			2. Supported Body Tilt			3. Unsupported Body Tilt		
	df	*MS*	*F*	*df*	*MS*	*F*	*df*	*MS*	*F*
*S*s	23	.0743	<1.00	11	.0331	1.08	23	.0104	<1.00
Sequence (S)	2	.0560	<1.00	2	.0440	1.43	2	.0030	<1.00
*S*s within S	21	.0761	—	9	.0307	—	21	.0111	—
Order	2	.0065	<1.00	2	.0105	1.35	2	.0020	<1.00
Tilt	2	.0820	3.40*	2	.0355	4.55*	2	.0365	5.37*
Square uniqueness	2	.0395	1.64	2	.0055	<1.00	2	.0005	<1.00
Residual	42	.0241	—	18	.0078	—	42	.0068	—
Total	71	.0419	—	35	.0174	—	71	.0085	—

* $P < .05$.

rotated in a vertical plane around an axis at the center of *S*'s body to three body positions: L, E, and R. The aesthesiometers were attached to a platform mounted on the box in a fixed position relative to *S*'s body. Twelve *S*s were tested. In Cond. 3 (Unsupported Head and Trunk Tilt) *S* stood and with head and trunk straight assumed three positions: head and trunk L, E, and R. A scale behind *S* assured that he maintained the required position. The aesthesiometers were mounted on a platform which was tilted in the same direction and amount as *S*'s head and trunk. Twenty-four *S*s were tested.

Results.—The analysis of variance for Cond. 1 head tilt (Table 1), shows that the F test for tilt is significant. In general the left palm is more sensitive than the right; however the relative differences between conditions is in accord with the hypothesis. The relative sensitivity of the left palm is greatest for head tilt right (−2.00 mm.), less for head erect (−1.54 mm.), and least for head tilt left (−0.25 mm.) (Table 2).

The analysis of variance for Cond. 2, supported body tilt (Table 1), shows that the F test for tilt is significant. For all conditions the right palm is more sensitive than the left palm; however, the relative differences between the conditions are in accord with the hypothesis. The relative sensitivity of the right palm is greatest for head tilt left (+4.33 mm.), less for head erect (+3.08 mm.), and least for head tilt right (+1.34 mm.) (Table 2).

The analysis of variance for Cond. 3, unsupported head and trunk tilt (Table 1), shows that the F test for tilt is significant. The differences between the conditions are in accord

TABLE 2

EXPERIMENT 1: MEAN TWO-POINT THRESHOLDS (MILLIMETERS) FOR RIGHT AND LEFT PALMS

Tilt Condition	*N*	20° Left			Erect			20° Right		
		Left Palm	Right Palm	Diff. LP-RP	Left Palm	Right Palm	Diff. LP-RP	Left Palm	Right Palm	Diff. LP-RP
1. Head tilt	24	8.83	9.08	−0.25	8.92	10.46	−1.54	8.71	10.71	−2.00
2. Supported body tilt	12	11.16	6.83	+4.33	10.75	7.67	+3.08	9.92	8.58	+1.34
3. Unsupported body tilt	24	8.58	8.33	+0.25	9.17	9.88	−0.71	8.25	9.58	−1.33

with the hypothesis. The relative sensitivity of the right palm is greatest for head tilt left (+0.25 mm.), less for head erect (−0.71 mm.), and least for right head tilt (−1.33 mm.) (Table 2).

EXPERIMENT II

Generalizing from earlier findings showing a concomitant reorganization in visual space with a redistribution of visual sensitivity, the following hypothesis was formulated: tilt left (right) shifts the apparent median plane to the right (left). The same sets of postural conditions as in Exp. 1 are used here: head tilt, supported body tilt, and unsupported head and trunk tilt.

Method.—Each *S* made a tactual-kinaesthetic adjustment of a perpendicular rod to a position which appeared straight ahead of the body (in the apparent median plane). Independent groups of *S*s were used for each set of postural conditions. As in Exp. I, each set of conditions consisted of three positions: tilt 20° left (L), erect (E), and tilt 20° right (R). For each of these positions two starting positions of the rod were used: 30 cm. left and 30 cm. right of the center of the body (the objective median plane). Thus there was a total of six trials which were arranged in a 6 × 6 replicated latin square design.

The perpendicular rod could be moved left and right in a horizontal track 6 in. in front of *S*. Location of the rod is measured in centimeters with respect to the objective median plane. With *S*'s eyes closed, his fingers are placed on the perpendicular rod. Instructions are to move the rod to the left or right until it is located straight ahead of the body.

The position of the apparent median plane in terms of deviations from the objective median plane is utilized. A plus deviation indicates a shift of the apparent median plane to the right and a minus deviation indicates a shift to the left.

Condition 1 (Head Tilt) was the same as in Exp. I, with 18 *S*s. Condition 2 (Supported Body Tilt) was the same as in Exp. I, with 24 *S*s. Condition 3 (Unsupported Head and Trunk Tilt) was also the same as in Exp. I, with 24 *S*s.

Results.—The analysis of variance for Cond. 1, head tilt, shows that tilt is a significant source of variance (Table 3). The differences between conditions are in accord with the hypothesis. For head erect the apparent median plane is +0.06 cm. from the objective median plane. With left tilt, the apparent median plane shifts relatively to the right (+0.82 cm.), and with right tilt, the

TABLE 3

EXPERIMENT II: ANALYSES OF VARIANCE FOR LOCATION OF APPARENT STRAIGHT AHEAD

Source	1. Head Tilt			2. Supported Body Tilt			3. Unsupported Body Tilt		
	df	*MS*	*F*	*df*	*MS*	*F*	*df*	*MS*	*F*
*S*s	17	8.815	1.08	23	21.279	1.06	23	15.457	1.01
Sequence (S)	5	10.294	1.25	5	25.707	1.28	5	15.955	1.04
*S*s within S	12	8.199	—	18	20.049	—	18	15.318	—
Order	5	2.694	<1.00	5	26.601	2.32	5	20.864	1.97
Conditions	5	55.772	14.99*	5	175.887	15.33*	5	200.236	18.89*
Tilt (T)	2	12.380	3.33*	2	69.797	6.08*	2	98.461	9.29*
Starting position (SP)	1	250.560	67.33*	1	725.400	63.24*	1	641.356	60.51*
T × SP	2	1.770	<1.00	2	7.219	<1.00	2	81.452	7.68*
Square uniqueness	20	7.938	2.13	20	15.522	1.35	20	13.366	1.26
Residual	60	3.721	—	90	11.471	—	90	10.600	—
Total	107	7.703	—	143	19.892	—	143	18.757	—

* $P < .05$.

TABLE 4

EXPERIMENT II: MEAN LOCATION OF APPARENT STRAIGHT AHEAD (CENTIMETERS)

Tilt Condition	N	20° Left		Erect		20° Right	
		App.	Diff. Phys.-App.	App.	Diff. Phys.-App.	App.	Diff. Phys.-App.
1. Head	18	49.18	+0.82	49.94	+0.06	50.34	−0.34
2. Supported body	24	47.51	+2.49	48.57	+1.43	49.92	+0.08
3. Unsupported body	24	46.98	+3.02	48.33	+1.67	50.43	−0.43

Note.—Physical straight ahead = 50; less than 50 is right and greater than 50 is left of physical straight ahead.

apparent median plane shifts relatively to the left (−0.34 cm.) (Table 4).

The analysis of variance for Cond. 2, supported body tilt, shows that tilt is a significant source of variance (Table 3). Even though for all conditions the apparent median plane is located to the right of the objective median plane, the relative differences between these conditions are in accord with the hypothesis. For body erect the apparent straight ahead is +1.43 cm. from the objective median plane. With left tilt, the apparent straight ahead shifts relatively to the right (+2.49 cm.), and with right tilt, the apparent straight ahead shifts relatively to the left (0.08 cm.) (Table 4).

The analysis of variance for Cond. 3, unsupported head and trunk tilt, shows that tilt is a significant source of variance (Table 3). Differences between conditions are in accord with the hypothesis. For head and trunk erect the mean location of the apparent straight ahead is +1.67 cm. from the objective median plane. With left tilt the apparent straight ahead shifts relatively to the right (+3.02 cm.), and with right tilt the apparent straight ahead shifts relatively to the left (−0.43 cm.) (Table 4).

DISCUSSION

With respect to the first aim of these studies, the findings for normal adults are in accord with previous findings for cerebellar patients (Halpern & Kugelmas, 1957): the distribution of tactual sensitivity depends on the postural status of the organism. Further, we can say that the direction of postural change, body tilt left or right, has a systematic effect on the distribution of tactual sensitivity, viz., sensitivity increases right or left, respectively. On this basis we may propose that a comprehensive explanation of tactual sensitivity must take into account postural factors as well as the tactual afferent system.

In regard to our second aim, the present findings permit generalization of the previously found relationship in the visual modality to the tactual modality: there are concomitant changes in the distribution of sensitivity and the organization of space depending on postural changes.

A variety of studies conducted by other investigators may be adduced in support of the generalization offered here that postural changes also affect the distribution of tactual sensitivity and the tactual-kinaesthetic organization space. Concerning sensitivity, both temporary and enduring changes due to postural factors have been reported. As to temporary changes, it has been reported (Frank, 1957) that when the hand and fingers are resting inactive the third finger is the most sensitive but when the hand and fingers are actively searching the forefinger is the most sensitive. Enduring changes, due to amputation and brain injury have been reported respectively by Teuber, Krieger, and Bender (1949); Haber (1955); and Halpern and Kugelmas (1957). Teuber and Haber have found that the sensitivity on all remaining parts of an amputated arm is proportionally greater than on the homologous parts of the amputee's intact arm. Halpern and Kugelmas report that cerebellar patients who manifest balanced asymmetrical posture to one side show an increase of tactual sensitivity on the opposite side of the body. These changes in tactual sensitivity—both transient and enduring, point to the role of postural factors in the distribution of sensitivity: a change in sensitivity with a shift from inactivity to activity of the hand; an increase in sensitivity in the stump of the amputee's arm when it acquires the function of operating as a manipulative organ; finally, the change in tactual sensitivity with a shift from a symmetrically balanced to an asymmetrically balanced stance.

Temporary and enduring changes have also been reported in the tactual-kinaesthetic organization of space. In regard to temporary changes, Werner, Wapner, and Bruell (1953) have shown that head turning systematically shifts the direction of pointing when subjects are asked to point to the straight-ahead.[2] Goldstein (1936) has reported enduring changes for cerebellar patients who adopt abnormal postures: they shift the locus of the middle of their bodies depending on the side of lesion. Halpern (1951) has also found that patients with cerebellar and vestibular disturbance shift the location of the middle of their bodies.

The work of others just reviewed, an earlier set of studies in the visual modality, and the experiments presented here, all point to the fact that with changes in postural status there are changes *both* in the distribution of sensitivity and in the organization of space. We may, however, go one step further. Our previous studies in vision and the present studies in the tactual modality point to systematic directional shifts in the spatial distribution of sensitivity and in the organization of space; that is, with an increase of sensitivity to one side, there is a shift of the apparent median plane to that side. It seems therefore plausible on this basis to postulate a bond between these two shifts—in the distribution of sensitivity and the organization of space—occurring with changes in postural status. As is well known, some theorists such as Goldstein (1939, pp. 53–55) would view this relationship in terms of its value for the adaptation of the organism. Our eventual aim, however, is to account for this bond between sensitivity and space organization in terms of a mechanism which permits the reciprocal interaction of sensory and postural factors.

SUMMARY

Experiments have been reported which show the effect of postural tilts left and right on (*a*) the distribution of tactual sensitivity as measured by the two-point threshold of the palms, and on (*b*) the organization of tactual-kinaesthetic space as measured by the location of the tactual-kinaesthetic straight-ahead. Tilt left increases sensitivity of the right palm compared to the left palm, and also shifts the straight-ahead to the right. Tilt right increases sensitivity of the left palm compared to the right palm, and also shifts the straight-ahead to the left. These findings indicate the necessity of considering postural factors as well as the tactual afferent system in a comprehensive explanation of sensitivity. Further, these findings are viewed as extending the previously found relationship between the distribution of visual sensitivity and the organization of visual space depending on postural factors: that is, for both the visual and the tactual modalities there are systematic directional shifts depending on postural factors. This evidence makes it plausible to postulate an inner relationship between these two types of shifts—distribution of sensitivity and organization of space.

REFERENCES

BORING, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century, 1942.

---

[2] We may refer here also to related studies by Morant (1952) which demonstrate systematic shifts in the tactual-kinaesthetic median plane depending on labyrinthine stimulation due to rotation.

BRUELL, J. Visual egocentric localization, an experimental study. Unpublished doctoral dissertation, Clark University, 1953.

FRANK, L. K. Tactile communication, *Genet. psychol. Monogr.*, 1957, **56**, 209–256.

GIBSON, E. J. Improvements in perceptual judgments as a function of controlled practice or training. *Psychol. Bull.*, 1953, **50**, 401–432.

GOLDSTEIN, K. The function of the cerebellum from a clinical standpoint. *J. nerv. ment. Dis.*, 1936, **83**, 1–12.

GOLDSTEIN, K. *The organism.* New York: American Book., 1939.

HABER, W. B. Effects of loss of limb on sensory functions. *J. Psychol.*, 1955, **40**, 115–123.

HALPERN, L. The syndrome of sensory-motor induction in combined cerebellar and labyrinthine injury. *J. nerv. ment. Dis.*, 1951, **114**, 137–161.

HALPERN, L., & KUGELMAS, S. The variability of tactile two-point discrimination in the sensory-motor induction syndrome. *Acta psychol., Amst.*, 1957, **13**, 162–172.

JENKINS, W. L. Somesthesis. In S. S. Stevens (Ed), *Handbook of experimental psychology.* New York: Wiley, 1951, Pp. 1172–1190.

MCFARLAND, J. H. The effect of asymmetrical eye torsion on perceptual clarity. Unpublished master's thesis, Clark University, 1957.

MCFARLAND, J. H., WERNER, H., & WAPNER, S. The effect of muscular involvement on sensitivity: Asymmetrical convergence on the distribution of visual sensitivity. *Amer. J. Psychol.*, 1960, **73**, 523–534.

MORANT, R. B. Factors influencing the apparent median plane under conditions of labyrinthian stimulation. Unpublished doctoral dissertation, Clark University, 1952.

TEUBER, H.-L., KRIEGER, H. P., & BENDER, M. B. Reorganization of sensory function in amputation stumps: Two-point discrimination. *Fed. Proc.*, 1949, **8**, 156.

WEBER, E. H. Der Tastsinn und das Bemeingefuhl. In R. Wagner (Ed), *Handworterbuch der Physiologie III.* Braunschweig: Vieweg, 1846, Pp. 481–588.

WERNER, H., WAPNER, S., & BRUELL, J. Experiments on sensory-tonic field theory of perception: VI. Effect of position of head, eyes, and of object on position of the apparent median plane. *J. exp. Psychol.*, 1953, **46**, 293–299.

(Received November 16, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 155–159

ACQUISITION AND TRANSFER IN PARTIAL REINFORCEMENT

E. J. CAPALDI AND MONTE G. SENKO[1]

University of Texas

According to the aftereffects hypothesis, reinforcement (or nonreinforcement) on a particular trial serves to give rise to distinctive stimuli which may persist until the following trial and become conditioned to responses. Although the hypothesis was originally formulated to deal with extinction (Sheffield, 1949), it may also be employed in connection with acquisition and transfer phenomena. Experimental results relevant to these latter aspects of the hypothesis have recently become available (Bloom & Capaldi, 1961). However, these data are quite limited; neither are they as critically related to the hypothesis as seems desirable, nor do they exclude alternative interpretations with a degree of sufficiency which seems necessary. The major purpose of this investigation was to test specific deductions available from an aftereffects hypothesis with regard to acquisition and transfer using regular patterns of reinforcement. There were certain lesser purposes as well, which will be discussed as occasion arises.

Consider a reinforcement pattern in which two consecutive nonreinforced trials are consistently followed by a reinforced trial (regular 33%). In this case the aftereffects of nonreinforcement are equally often followed by reinforcement and nonreinforcement, while the aftereffects of reinforcement are followed only by nonreinforcement. Prediction from the hypothesis in this instance is clear. It is expected that completely appropriate pattern running, such as develops in the single alternation situation (SA) will not occur (Bloom & Capaldi, 1961; Capaldi, 1958; Tyler, Wortz, & Bitterman, 1953). Rather, the hypothesis predicts that the rat will run appropriately (relatively rapidly) on the reinforced trial (R), appropriately (relatively slowly) on the nonreinforced trial following the reinforced one (N_1), and inappropriately (relatively rapidly) on the succeeding nonreinforced trial (N_2). It is further expected that running times on Trials R and N_2 should be practically identical.

Assuming that the rat responds in the manner outlined above, further predictions regarding transfer are available from the hypothesis. As indicated, the hypothesis holds that during acquisition in the regular 33% situation *S* will form the habits of running rapidly following nonreinforcement on the one hand, and slowly following reinforcement on the other. But, according to the hypothesis, these are precisely the habits which are formed in the SA situation. Therefore, upon transfer, under the conditions outlined above, it is expected that immediate, consistent, and completely appropriate SA pattern running will occur.

METHOD

Subjects.—The *S*s, approximately 95 days old at the beginning of the experiment, were 31 male Wistar strain rats purchased from Barlan Industries, Cumberland, Indiana.

Apparatus.—The apparatus was a straight-alley runway enclosed by walls, with a battery of five goal boxes mounted on runners.

---

[1] Now at the University of Wisconsin.

Each goal box could be positioned behind a one-way, gray, swinging door which separated goal box from alley. Goal boxes not in line with the runway were neither visible nor accessible to *S* from the alley. Placing *S* down at the start point of the alley depressed an 8 × 3⅝ in. platform supported by a microswitch and simultaneously activated a timer. Upon touching the 3⅗ × 6 in. swinging door at the alley's end, contacts were broken and the timer, which measured in .001 sec., was stopped. The elapsed time was recorded.

The dimensions of the alley, which was painted gray, were 66 × 3¾ × 6 in. The colors of the goal boxes from left to right were gray, gray, white, black, and white. Reinforcement occurred in one gray or white box and nonreinforcement in the other gray or white box to control for the presence of food particles on nonreinforced trials. The inside dimensions of each goal box were 11½×4¼×6 in. A hinged, hardware-cloth lid covered the entire battery of goal boxes. Illumination was provided by a fluorescent lamp placed 5 ft. above the starting point.

Procedure.—On Days 1 through 5 *Ss* were handled and fed wet mash 1 hr/day. On Day 6 each *S* was given six preliminary trials, three from 1 ft. in front of the goal box and three from a point about halfway in the alley. On this day the swinging door was raised and food was present in the goal box. On Day 7 each *S* received six runs from the start point of the alley with swinging door down and food in the goal box. Throughout all phases of the experiment *S* was confined to the goal box for 20 sec. following each run, the intertrial interval was 20 sec., and *S* was fed for 1 hr. in the home cage immediately following the daily runs.

Experimental training was begun on Day 8. All *Ss* were given 12 trials per day and 33% reinforcement. For Group IR-G ($N=8$) reinforcement was random and both reinforcement and nonreinforcement occurred in a gray goal box. Group R-G ($N = 8$) was comparable in all respects save that reinforcement was regular—the initial trial of the day being nonreinforced, the third reinforced, and so on. Finally, Group R-WBW ($N=15$) was likewise regularly reinforced; however, on trials preceding reinforcement *S* found a black goal box and on trials preceding nonreinforcement, a white goal box. The *S* was thus both nonreinforced (Trials 1, 4, 7, and 10) and reinforced (Trials 3, 6, 9, and 12) in a white goal box.

On Day 15 all three groups were transferred to an SA pattern. Both reinforced and nonreinforced trials occurred as before in gray goal boxes. Group R-G received 7 days of SA training, Group IR-G received 9 days of such training. Group R-WBW was trained as before on the regular 33% pattern for an additional 7 days, i.e., a total of 21 days. The intertrial interval, etc., were as before.

RESULTS

In Fig. 1 are presented the log medians of median running times of Group R-WBW on Trials N_1, N_2, and R, on each of the 21 days of acquisition training. On the initial days *Ss* ran relatively slowly on R trials and increasingly more rapidly on Trials N_2 and N_1, respectively. The Day 1 and Day 2 differences (analysis using log medians) were significant ($F = 10.43$ and 8.36, $df = 2/28$ each case, $P < .01$ each case). The tendency for slow running on N_1 trials began on Day 6 and increased consistently thereafter until Day 11, when it seems to have reached full strength. As can be seen, Trial N_2 and R running times are practically identical from about Day 6 onward. While the differences shown in Fig. 1 on Day 6 appear small, they are nevertheless significant ($F = 8.95$, $df = 2/28$, $P < .01$). A separate analysis of N_2 and R times failed to reveal significant differences ($F<1$), thus differences in the overall analysis must be due to slower running on N_1 trials. Actually, on Day 6, 13 of the

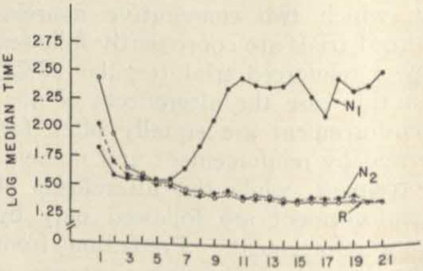


FIG. 1. Log median of median running times of Group R-WBW for each of the acquisition days.

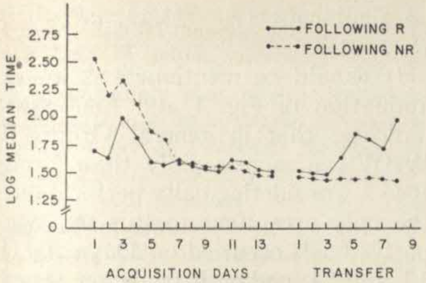


FIG. 2. Log median of median running times of Group IR-G for each of the acquisition and transfer days.

15 *S*s ran more slowly on N_1 trials than on either N_2 or R trials.

Since, later in training, *S*s ran equally rapidly on both N_2 and R trials it appears that the presence of the white goal box on all trials preceding nonreinforcement did not serve as a particularly effective cue from the point of view of pattern running. If it had, that is, if color of goal box gave rise to distinctive aftereffects, as apparently does food or its absence, then N_2 running times presumably should have been slower than those on R trials.

In Fig. 2 the log median of median running times of Group IR-G on trials following reinforcement and on trials following nonreinforcement on each of the acquisition and transfer days are presented. On each of the initial 6 days of acquisition training, *S*s ran more rapidly on trials following reinforcement. These daily differences, except for those on Day 5, are significant from beyond the .05 to beyond the .01 level ($F = 14.30$, 19.02, 6.33, 15.30, 4.20, and 7.13 for Days 1 through 6 respectively; $df = 1/7$ in each case). On Days 7 through 10, running speeds are about equal on both types of trials; however, from Day 11 onward *S*s show a slight tendency to run more slowly on trials following reinforcement. The log of the sums of the median running speeds for each *S* on trials following reinforcement and on trials following nonreinforcement on Days 11 through 14 were analyzed. The differences very closely approach but do not reach significance at the .05 level ($F = 5.37$; needed at the .05 level for 1/7 *df*, an F of 5.59). The numbers of transitions of each type on the initial 10 days of training were R-R, 12; R-N, 24; N-R, 28; and N-N, 46. Thus, the percentage of reinforced trials followed by a reinforced trial is 33.3 and the percentage of nonreinforced trials followed by a reinforced trial is 37.8. It seems fair, if not necessary, to conclude that the differential running was due to the slight differences in percentage indicated above since, when nonreinforced trials are followed by an exactly equal number of reinforced and nonreinforced trials, identical running times on N_2 and R trials result. As will be seen, this was true not only for Group R-WBW but for Group R-G as well. That such small differences may result in differential running seems surprising.

The tendency to run more slowly following reinforcement persisted into the transfer stage. Analysis of log median differences on the initial day of SA training yielded an F of 4.75 which, for 1/7 *df*, is not significant. Despite the fact that under the SA pattern R trials are consistently followed by N trials, the already slightly developed tendency to run slowly following R trials did not increase in strength until Day 4 of training. By this day the differences reached significance ($F = 6.49$, $df = 1/7$, $P < .05$).

In Fig. 3 are presented the log median of median running times of Group R-G during acquisition and transfer. In the acquisition stage these times are presented on the basis of R, N_1, and N_2 trials, and in the transfer stage, on the basis of trials

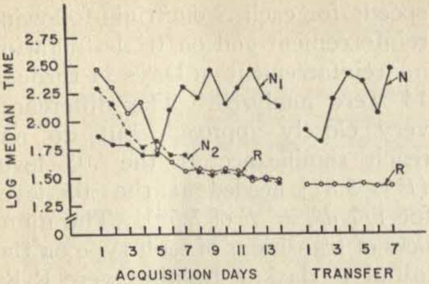


FIG. 3. Log median of median running times of Group R-G for each of the acquisition and transfer days.

following reinforcement and trials following nonreinforcement. On each of the initial 4 days of acquisition, as before, running times were most rapid on N_1 trials, less rapid on N_2 trials, and least rapid on R trials. Only the differences on the initial day were significant ($F = 22.16$, $df = 2/14$, $P < .01$). Running times on N_2 and R trials are practically identical from Day 8 onward, and the tendency for slow running began on Day 6 and appears to have reached full strength on Day 9. As in the case of Group R-WBW, significantly slower running on N_1 trials appeared initially on Day 6 ($F = 4.93$, $df = 2/14$, $P < .05$).

When shifted to the SA pattern, *S*s showed a tendency to run more slowly on nonreinforced trials, i.e., on trials following reinforcement. However, Fig. 3 does not show an important aspect of these data. The least rapid of the initial day's reinforced trial running times was shorter than the shortest of that day's nonreinforced trial running times. In other words, on the initial day, a saw-toothed, SA pattern running curve was obtained when the data were plotted on a trial by trial basis. An analysis of the log median times on reinforced vs. nonreinforced trials on the initial day showed them to be significant ($F = 22.42$, $df = 1/7$, $P < .01$).

It should be mentioned as an examination of Fig. 1 and Fig. 3 will indicate, that in general Group R-WBW ran more rapidly than Group R-G. Considering daily performance, the only exceptions to this tendency on N_1 trials occurred on Days 10, 11, 13, and 14; and on R trials, on Day 1. No exceptions occurred considering N_2 trials. Daily differences between the groups were tested for each type of trial for each of the initial 8 days. In each case, $df = 1/21$ and, with a single exception, differences reported were significant beyond the .05 level. Group R-WBW ran significantly more rapidly than Group R-G on R trials on Day 3 ($F = 5.72$), Day 4 ($F = 7.11$), and Day 5 ($F = 4.96$); on N_1 trials on Day 4 ($F = 11.46$, $P < .01$); and on N_2 trials on Day 3 ($F = 4.36$), Day 4 ($F = 4.48$), Day 5 ($F = 6.33$), Day 6 ($F = 6.17$), and Day 7 ($F = 7.24$). Since the greatest effect seems to have been on N_2 and R trials, i.e., on trials following nonreinforcement, rather than on N_1 trials, which followed reinforcement, it seems that Group R-G was more susceptible to relatively slow running following nonreinforcement.

DISCUSSION

The predictions of major concern have been consistently borne out by the data and need little comment. What does appear to be in need of comment, however, is the relevance of the present results for an alternative hypothesis which was the original source of and has been consistently applied to pattern learning experiments of the present type. In addition, this hypothesis is compatible with the results of most investigations which have employed irregular patterns of reinforcement.

According to the serial patterning version of the discrimination hypothesis, partial reinforcement cannot be understood on a trial by trial basis but requires, rather, a gestalt-like level of analysis which considers that the rat responds to the entire series of events, i.e., learns a pattern per se (Capaldi, 1958; Tyler et al., 1953; Wodinsky & Bitterman, 1959). With reference to the initial proposition, that a trial by trial analysis is theoretically improper, the experimental results would seem to speak for themselves. That the rat learns a pattern per se, the second of the proposals entertained by the hypothesis, seems quite unlikely from these results. Actually, two somewhat distinct criticisms are implied here. First, a trial by trial aftereffects analysis is able to explain that which the serial patterning hypothesis purports to explain but in fact merely accepts without explanation, e.g., pattern running in the SA situation. Moreover, in specific situations, the aftereffects hypothesis yields predictions utterly unavailable to the patterning hypothesis. As far as can be determined, a patterning view cannot be made to indicate whether the rat will respond completely appropriately in the presence of a regular 33% pattern and, if not, in what specific respects response will be inappropriate. Nor do specific predictions regarding transfer seem available from the patterning hypothesis.

The successes and failures of a discrimination hypothesis do not appear difficult to understand. That hypothesis, in common with a pure form of the aftereffects hypothesis, holds that resistance to extinction and amount of stimulus change are inversely related. Apparently, a simple examination of reinforcement patterns, despite a lack of knowledge of specific mechanisms, is adequate in most instances to determine which of two or more patterns most differs from the extinction situation, i.e., zero reinforcement. While this procedure has encountered some success as regards extinction, it appears less adequate in the case of transfer and entirely inadequate with reference to acquisition. As the present results have shown, a knowledge of specific mechanisms is required to achieve prediction in these latter instances.

SUMMARY

Three groups of rats were trained to traverse a runway under conditions of 33% reinforcement. Two of the groups were regularly reinforced. One of the regular groups was both reinforced and nonreinforced in a gray goal box. The other regular group was reinforced in a white goal box and nonreinforced in a white and in a black goal box. The third group was irregularly reinforced in a gray goal box. As predicted from an aftereffects hypothesis, the regular groups ran slowly following reinforcement and rapidly on the succeeding two trials. The regular gray group was transferred to a single alternation (SA) pattern of reinforcement. Immediate and consistent SA pattern running occurred. Color of goal box did not appear to be a factor in giving rise to distinctive aftereffects.

It was suggested that certain of the results could not be accounted for by the serial patterning version of the discrimination hypothesis.

REFERENCES

BLOOM, J., & CAPALDI, E. J. The behavior of rats in relation to complex patterns of partial reinforcement. *J. comp. physiol. Psychol.*, 1961, **54**, 261–265.

CAPALDI, E. J. The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, **51**, 367–371.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

TYLER, D. W., WORTZ, E. C., & BITTERMAN, M. E. The effect of random and alternating partial reinforcement on resistance to extinction in the rat. *Amer. J. Psychol.*, 1953, **66**, 57–65.

WODINSKY, J., & BITTERMAN, M. E. Partial reinforcement in the fish. *Amer. J. Psychol.*, 1959, **72**, 184–199.

(Received November 17, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 160–166

AMOUNT, DELAY, AND POSITION OF DELAY OF REINFORCEMENT AS PARAMETERS OF RUNWAY PERFORMANCE[1]

EUGENE R. WIST[2]

University of Missouri

The introduction of delay at some point in a sequence of behavior can exercise its effect in two places, on the responses preceding the point in the response chain where delay is introduced and on the responses following this point.

The effect of delay on the responses preceding the point of delay is reported as a gradual decrement in response strength over a series of trials to a new asymptote (Hull, 1952; Spence, 1956). Furthermore a gradient of response strength, the "goal gradient," is supposed to exist among the members of the response chain terminating in reinforcement (Hull, 1934). Thus the new asymptote after delay introduction should be some function of the point in the chain at which delay is introduced.

The effect of delay on the responses following the point of delay has been reported both as incremental (e.g., Bernstein, 1954) and decremental (e.g., Collier & Marx, 1957). Again if the effect, whatever it is, is proportional to the predelay response strength, then its magnitude would be some function of the point in the chain at which delay is introduced.

The present experiment attempts to make a direct examination of these conjectures by means of two, independent, presumed manipulators of response strength. First, the point in the response chain at which delay is introduced was varied. If a within-chain gradient exists, this should result in variations in response strength at different points in the chain. Second, the magnitude of reinforcement was varied to produce two degrees of response strength. Combinations of these two variables and different delay durations made possible a comparison of the effects of the two means of producing variation in response strength on pre- and postdelay running gradients. Furthermore, quantitative measurement and control of consummatory behavior made possible a determination of some of the relationships between end-box events and running gradients.

METHOD

Experimental design.—Two reinforcement concentrations (8% and 32% sucrose solutions), two delay intervals (3 and 45 sec.), and three positions of delay (2, 6, and 10 ft. from the goal box) were factorially combined. Two additional 8% and 32% control groups were run without delay.

Subjects.—The *S*s were 98 female albino rats with a mean age of approximately 4 mo.;

---

[1] This experiment represents a portion of a dissertation submitted to the Psychology Department of the University of Missouri in partial fulfillment of the requirements for the PhD degree. The author wishes to express his indebtedness to George Collier, and to Melvin H. Marx, his advisor, for their advice and encouragement. A report on a portion of these data was made at the 1960 Midwestern Psychological Association meetings in St. Louis.

This investigation was supported in part by research grants, M-817 and M-3328, from the National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.

[2] Now at the United States Army Medical Research Laboratory, Fort Knox, Kentucky.

53 *S*s were of the Holtzman strain, 45 were from the colony of the Department of Psychology, University of Missouri. Two rats were discarded for failure to run or drink, 1 because of illness, 4 because of injury in the apparatus, and 7 to equalize *N*s at 6 per cell. All *S*s were experimentally naive, and were assigned at random to groups.

Apparatus.—The apparatus was an 18-ft. straight runway housed in a light-tight, sound-deadened hull. The start box, stem, and goal box were 1½ ft., 14 ft., and 2½ ft. long, respectively. Two guillotine doors separated these boxes from the stem. The delay box consisted of a pair of movable guillotine doors separated by 2 ft. and mounted on a frame. The rear door was operated by *S*'s entry into the chamber, the transparent front door by means of a Hunter decade timer. Similarly the end-box door was operated by *S*'s entry.

The runway floor consisted of nine 2-ft. sections of angle aluminum. The start-box door was located 6 in. short of the end of the first section. The goal-box door was located 6 in. short of the beginning of the last section. Similarly, the position of the delay box was arranged so that the front and rear doors were located 6 in. before the ends of the sections they isolated.

The runway was automatic. The start-box door opened and the first clock was started 3 sec. after the start-box lid was lowered. When *S*s reached the end of the first 2-ft. section, the first clock stopped and the second started. When the third section was reached, the second clock stopped and the third started, etc.

Solutions were delivered by two 50-ml. burettes to each of which a drinking tube with a 4-mm. opening had been annexed. These tubes were ½ in. behind a ½-in. square opening in a panel located 6 in. from the end of the goal box. A 7-w. light 5 in. behind this panel illuminated the goal box and runway. Contact with the tube activated an electronic relay. Number of licks, duration of tube contact, and volume consumed were recorded.

The sucrose solutions were mixtures of commercial sugar and distilled water made 24 hr. before their use and maintained at room temperature.

A four compartment box simulating the goal box of the runway was used as a drinker trainer.

Pretraining.—Twenty-eight days before the first runway trial, *S*s were placed on a 22-hr. food deprivation schedule. On Day 13 the feeding period was reduced to 1 hr. per day and all *S*s were placed on a table top in groups of 6 and allowed to explore for 15 min. during which time *E* picked up and handled each *S* for two 1-min. periods. On Day 17 each *S* received 5 min. of drinking in the drinking trainer. Half of the *S*s received an 8%, and half a 32% sucrose solution. On Day 18, this period was reduced to 3 min. On Days 19 through 25, *S*s received 10 min. of exploration on top of the runway followed by a 10-min. adaptation period in the runway. The floor of the runway was strewn with .45 mg. Noyes food pellets. On Day 26 *S*s were again placed in the drinking trainer for a 3-min. period of drinking. On Days 27 and 28 of deprivation, *S*s were individually placed in the goal box until a total of 15 sec. of tube contact was recorded.

Runway trials.—The *S*s were given one trial per day for 52 days. For the first 26 days, no delay was administered, but the delay-box doors were in operation at the appropriate position in the runway allowing *S*s to adapt to the auditory, visual, and tactual cues resulting from their operation. For the last 26 trials delay was administered on each trial at the appropriate positions. The *S* was allowed a 20-sec. drinking period in the goal box after which it was removed and returned to its home cage.

RESULTS AND DISCUSSION

Three 5-trial blocks were analyzed. The first block, Trials 22–26, consisted of the last 5 trials before the introduction of delay. The second block, Trials 27–31, consisted of the first 5 delay trials. The last block, Trials 48–52, consisted of the last 5 trials. Analyses of variance (Lindquist Type I and Type VI) were performed on both runway and consummatory behavior measures. Table 1 presents the major significant results of these analyses. The complete analyses may be found elsewhere (Wist, 1960). Alexander (1946) tests of trend were performed on the 15 trials preceding the delay introduction for each concentration group independently. None of the *F*s was significant, suggesting that running speed was asymptotic at the time of delay introduction.

TABLE 1

Mixed Analyses of Variance Performed on Running and Consummatory Behavior Measures

Measure	Source[a]	*F*	*df*
Starting speed, Start box (Fig. 2A)	C	35.31**	1/70
	C × B	7.68**	1/140
	D × B	18.32**	2/140
Running speeds, Predelay box segment (Fig. 2B)	C	21.98**	1/70
	D × B	26.91**	2/140
	P (Trials 22–26)	22.03	1/60
Starting speeds, Delay box (Fig. 2C)	C	4.58*	1/70
	B	28.57**	2/140
	Delay	30.22**	1/70
	B	32.03**	1/70
	D × B	6.21**	1/70
Running speeds, Postdelay box segment (Fig. 2D)	C	25.13**	1/70
	B	65.16*	2/140
	D (Trials 27–31)	17.27**	1/60
	B	78.35**	2/140
	D × B	14.71**	2/140
	P (Trials 22–26)	5.71*	1/60
Number of licks (Fig. 3A)	C (Trials 22–26)	5.02*	1/60
Lick rate (Fig. 3D)	C (Trials 22–26)	9.29**	1/60
Ingestion rate (Fig. 3E)	C × B	3.63*	2/140

Note.—Only major significant results are included in this table.

[a] C = Concentration, D = Delay, B = Blocks, P = Position.

* $P < .05$.

** $P < .01$.

Five of the results of the present experiment are of special significance.

First, the "classical" gradient of increased speed in the direction of the goal was not obtained (Hull, 1943). Instead, there was an initial burst of speed followed by a slight decrement as the goal was approached, a finding typical of this runway (Knarr, 1959; Knarr & Collier, 1960). Figure 1 shows the running gradients for the control groups for the 5-trial blocks immediately preceding and following the introduction of delay for the experimental groups. The experimental group gradients were similar in form immediately before delay introduction (Trials 22–26). This figure also indicates that the operation of the goal-box doors without delay had no appreciable effect on running speed (Trials 27–31).

The discrepancy between the data of the present study and those which have provided evidence for the goal gradient (Hull, 1934) would seem to be due to the fact that external stimuli were minimal in the runway employed. It would seem that this goal gradient is obtained when runway stimuli can acquire control over units of the response chain or occasion competing behavior (Pereboom, 1958). Control over the chain by terminal reinforcing and discriminative stimuli produces running gradients similar to those of the present study (Collier, Knarr, & Marx, 1961).

The above failure to obtain a goal gradient is reflected in the lack of a differential effect of position of delay of reinforcement on running speeds. The two significant *F*s for position (Trials 22–26) in Table 1 reflect respectively the initial burst of speed and the terminal decline of speed before delay introduction (see Fig. 2B and D).

Second, delay resulted in decremental changes in performance (see Fig. 2). This finding is contrary to the implications of "frustration-drive" models (e.g., Amsel, 1958;

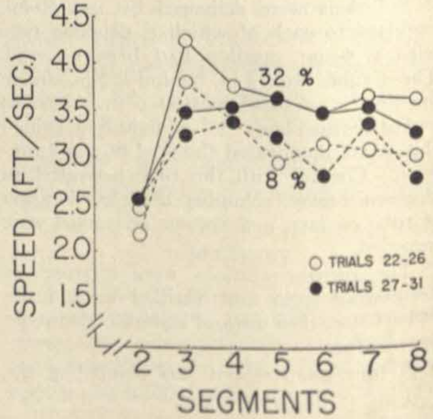


Fig. 1. Mean running speed for all segments over Trials 22–26 and Trials 27–31 for the control groups.

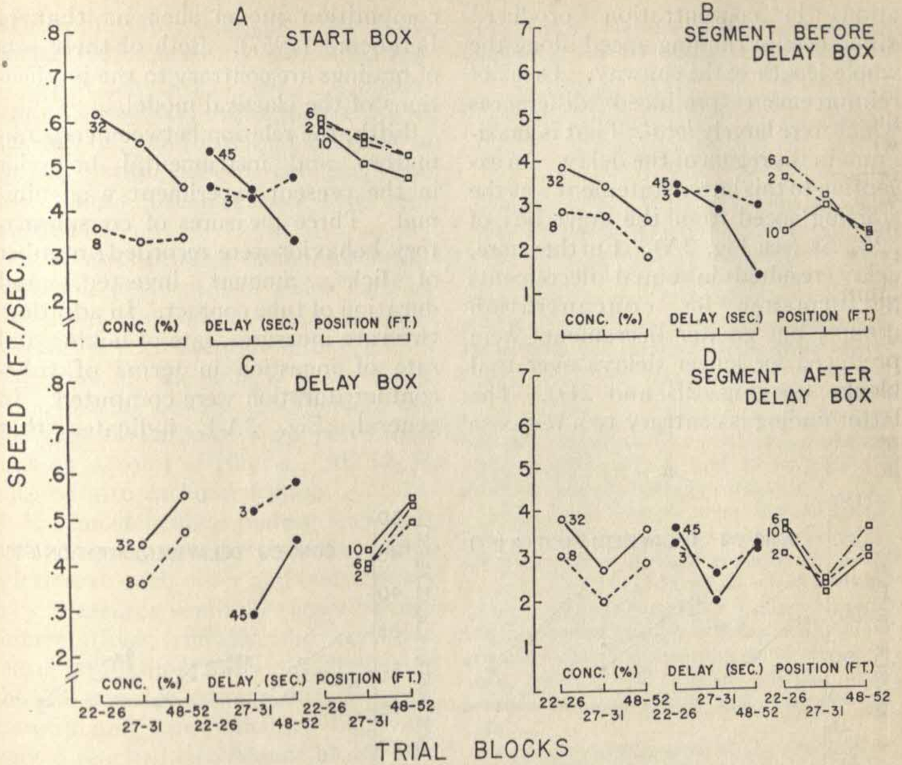


FIG. 2. Mean starting speeds for concentration, delay, and position groups from the start box (A) and delay box (C), and mean running speeds from the segment before the delay box (B) and the segment after the delay box (D) over trial blocks.

Brown & Farber, 1951). Since delay was not introduced until the second block of trials, there are only two points in Fig. 2C.

Third, delay had a differential effect on pre- and postdelay-box running speeds. Running speeds in the segments preceding the point of delay declined gradually to a new asymptote inversely proportional to delay duration. Running speeds in the segments following the delay box dropped immediately to a level proportional to the delay duration and then recovered gradually over trials to their original level (see Fig. 2B and D). These effects were significant (see Table 1, 2B, and 2D), and similar to those obtained by Collier and Marx (1957).

The effect on the responses preceding the point of delay is consistent with previous findings (Hull, 1952; Spence, 1956). According to the classical Hullian account delay weakens the habit strength (or reaction potential) of responses made more remote in time from the goal. According to this account the delay per se should not have affected postdelay-box running as it did, since the temporal relation of responses in this portion of the runway to the goal remained unchanged.

Fourth, delay and amount of reinforcement do not appear to be equivalent parameters of reinforcement. Variations in concentration produced differences in running speed along the whole length of the runway. Delay of reinforcement produced differences which were largely *local.* That is, maximum in the region of the delay. An exception to this latter statement was the starting speed from the start box of 32% *S*s (see Fig. 2A). Furthermore, delay resulted in equal decrements performance for concentration groups, but greater decrements were produced for longer delays over trial blocks (see Fig. 2B and 2D). The latter finding is contrary to a response competition model such as that of Pereboom (1957). Both of these sets of findings are contrary to the implications of the classical model.

Fifth, the relation between consummatory and instrumental behavior in the present experiment was minimal. Three measures of consummatory behavior were recorded: number of licks, amount ingested, and duration of tube contact. In addition, two rate measures, rate of licking and rate of ingestion in terms of tube-contact duration were computed. In general, Fig. 3A-E indicates that

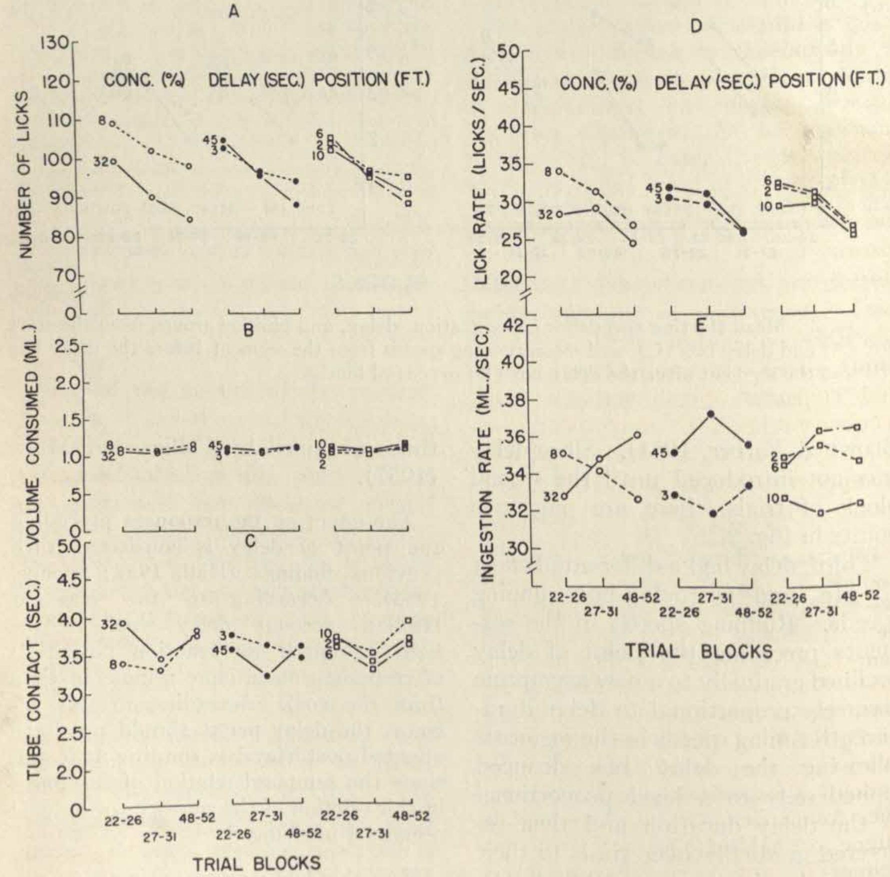


FIG. 3. Consummatory behavior measures for the three 5-trial blocks. (The lick rate and ingestion rate measures of 3D and 3E represent, respectively, the number of licks and amount ingested per second of tube contact.)

consummatory behavior was not greatly influenced by concentration, delay, or position of delay. Concentration, however, did influence initial lick number and rate and resulted in an interaction between concentration and blocks for ingestion rate (see Table 1).

A series of product-moment correlations was performed in which each of the nine dependent variables was correlated with every other one. This was done for the sums of each of the three 5-trial blocks upon which the analyses of variance were performed making a total of 108 *r*'s. All 72 *S*s entered into each correlation.

The most general finding was that while runway measures showed some relation to each other and consummatory measures similarly showed some interrelations, runway and consummatory measures were unrelated. Of the 60 possible correlations between runway and consummatory behavior only 3 reached significance at the .05 level. A complete summary of these relations is available elsewhere (Wist, 1960). This result supports the view that consummatory events have no necessary relation to instrumental behavior (Collier et al., 1961) and fails to support r_g models (e.g., Hull, 1952; Spence, 1956).

The above data, against the background of previous findings, suggests the following tentative formulation of the relationship between running gradients and variables affecting performance: "Running" occurs at rates determined by its consequences, such as the concentration of reinforcement, and it occurs everywhere at the same rate independent of its temporal or spatial distance from its climax. Factors such as fatigue or habituation may result in declines in rate over long distances. Factors such as obstacles and distracting stimuli may produce local disturbances in running, but do not manipulate its basic rate. Time and distance factors, important in the classical treatment of performance gradients do not readily account for the effects obtained.

SUMMARY

When delay interrupts a response chain terminating in reinforcement, the effect on the within-chain gradient is presumably some function of the response strength of the link delayed. In the present study response strength was manipulated in two ways: (*a*) by varying the position of the delay in the chain; (*b*) by varying the amount of terminal reinforcement. Two reinforcement concentrations (8% and 32% sucrose solutions), two delay intervals (3 and 45 sec.), and three delay positions (2, 6, and 10 ft. from the goal box) were factorially combined.

Five major results were discussed: (*a*) The failure to obtain the goal gradient; (*b*) the decremental changes in performance due to delay; (*c*) the differential effect of delay on pre- and postdelay-box running speeds; (*d*) the nonequivalence of delay and amount of reinforcement as parameters of reinforcement, and (*e*) the minimal relation found between consummatory and instrumental behavior.

A tentative formulation of the relationship between running gradients and reinforcement parameters was suggested.

REFERENCES

ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.

AMSEL, A. The role of frustrative non-reward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BERNSTEIN, B. B. Extinction as a function of frustration. Unpublished doctoral dissertation, University of Missouri, 1954.

BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.

COLLIER, G., KNARR, F. A., & MARX, M. H. Some relations between the intensive properties of the consummatory response and reinforcement. *J. exp. Psychol.*, 1961, **62**, 484–495.

COLLIER, G., & MARX, M. H. Frustration and performance. *Amer. Psychologist*, 1957, **12**, 458. (Abstract)

HULL, C. L. The rat's speed of locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, **17**, 393–422.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KNARR, F. A. Performance as a function of differential consummatory activity. Unpublished master's thesis, University of Missouri, 1959.

KNARR, F. A., & COLLIER, G. Taste and consummatory activity in amount of reinforcement effects. Paper read at Midwestern Psychological Association, St. Louis, April 1960.

PEREBOOM, A. C. A note on the Crespi effect. *Psychol. Rev.*, 1957, **64,** 263–265.

PEREBOOM, A. C. A goalless gradient. *J. exp. Psychol.*, 1958, **55,** 31–33.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WIST, E. R. Amount, delay, and position of delay of reinforcement as parameters of runway performance. Unpublished doctoral dissertation, University of Missouri, 1960.

(Received November 21, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 167–176

THE INFLUENCE OF EARLY EXPERIENCE ON THE FRUSTRATION EFFECT[1]

ABRAM AMSEL AND ELIZABETH C. PENICK
University of Toronto *Washington University*

This experiment was designed to test the influence of early experience and training on the frustration effect (FE) as defined by Amsel and others (Amsel & Hancock, 1957; Amsel & Roussel, 1952; Wagner, 1959). It arises, primarily, out of a difference in interpretation of the phenomenon. While some workers in this area (e.g., Amsel, 1958; Brown & Farber, 1951) have leaned to a nonassociative interpretation of the FE, others (e.g., Lawson & Marx, 1958; Marx, 1956) have suggested that an associative mechanism accounts for it equally well. While the two points of view are not incompatible—associative and nonassociative factors might certainly operate together in the FE—it seems desirable to initiate some experiments to determine the extent to which variations in early experience factors, which can only affect later behavior associatively, determine the extent of the FE, as it has usually been measured in the double-runway apparatus. The investigation involved three experimental variables.

Rationalizing the first experimental variable, the number of pups to the litter in which *S* was reared, it is reasonable to suppose that in a large litter (12 pups) greater activity is required in securing food than is required in a small litter (3 pups); observation of large and small litters in the act of feeding demonstrates this. The differential amount of activity required of each type of litter in relation to temporarily unavailable food could be ascribed to several factors: the amount of space around the mother, the availability of a functioning nipple, and the amount of food available. It seems reasonable to suppose, therefore, that in the period before weaning large-litter animals are more likely to be frustrated when hungry, and are more likely to be active in response to frustration than animals raised in small litters, who are less likely to encounter factors inducing frustration in relation to feeding, and to learn vigorous activity in this connection.

The second experimental variable, conditioning of activity and inactivity to the frustrative presentation of food, was introduced after weaning, and was a further attempt to manipulate the early experience of *S*s. Some of the food-deprived *S*s were trained to vigorous activity in the presence of inaccessible food, while others were given food only after a period of relative inactivity. It appears reasonable to suppose, on the basis of a purely associative interpretation, that those *S*s trained to be inactive in the presence of inaccessible food would manifest a frustration effect, at later testing, of a lesser magnitude than those *S*s trained to be active in the presence of inaccessible food.

The third experimental variable was introduced during the test period in the double runway on nonreward trials. Half of the *S*s found an empty food cup in G_1 on nonreward trials, whereas the other half found food present but inaccessible (a wire covering allowed *S* to see and smell the food but not eat it). This variable may be rationalized as follows: food-not-present in G_1 is the usual frustration-defining situation; food-present-but-inaccessible in G_1 is a situation more similar to the frustrative feeding experiences of *S*s in the previous training stage, and it might be expected that there would be greater "transfer" of learned vigor or quiescence to this frustrative situation than to the food-absent situation.

---

[1] This experiment was performed at Tulane University and was supported by a grant, G-5527, from the National Science Foundation.

If the frustration effect is a result of, or is strongly affected by, learning which has occurred prior to the actual testing situation, then the experiment should support the following generalizations: *S*s which learned to respond to nonreward situations by increases in activity, particularly those from large litters, should when tested in the double-runway apparatus, particularly under the food-present-but-inaccessible condition, show the greatest frustration effect. Under the same test conditions, *S*s which learned early to respond to nonreward situations with a decrease in activity, particularly those from small litters, should show the smallest frustration effect.

METHOD

Subjects

The *S*s were 40 hooded rats from the colony maintained at Newcomb College, Tulane University. Two *S*s died just before testing. The *S*s were tested when approximately 107 days of age, although the experimental procedures began within a few days after their birth.

Apparatus and Procedure

The entire experiment was conducted in four stages: (*a*) Birth of *S*s to weaning; (*b*) isolation and activity-inactivity treatments; (*c*) preliminary adjustment to double runway, and test in activity cages under stimulus change conditions; (*d*) test in the double-runway apparatus. Each stage will be described separately both with respect to apparatus and procedure employed.

Stage 1.—This stage extended from birth of *S*s to weaning at approximately 31 days of age. Ten mother rats produced 58 pups. When approximately 4 days old, they were taken from their biological mothers and a split-litter technique was used to separate the pups into two groups of 24 each. Two litters of 12 pups each were formed from one group and eight litters of 3 pups each were formed from the other. Between the two litter-size conditions control for original litter differences was exercised as far as was possible. Of the 24 rats raised in a large litter, 20 were randomly selected when approximately 31 days of age for the large litter (L) condition. Of the 24 rats which were raised in a small litter, 20 were randomly selected for the small litter (S) condition.

Stage 2.—Stage 2 extended from weaning at Day 31 to Day 102. When *S*s were approximately 31 days of age, they were separated and put into individual living cages. The housing construction, composed of $6\frac{1}{2} \times 11 \times 7$ in. cages, was in a separate room, dimly and indirectly illuminated at all times by two 7½-w. bulbs. For 9 days *S*s were allowed to adapt to their home cages and were maintained on an ad libitum diet for food and water. The *S*s were then placed on 11½ hr. food deprivation with water present at all times. Four days were allowed for *S*s to adjust to the new feeding schedule and become adapted to the frustrative-feeding apparatus.

The frustrative-feeding apparatus was a $1 \times 1 \times 1$ ft. box of unpainted plywood with a clear glass covering. Placed 1 ft. above the box was a 15-w. bulb which made the apparatus much brighter than *S*s' home cages. A wall separated *E* and the home cages from the apparatus, and during the frustrative-feeding treatments *E* observed *S*s' behavior through a 2-in. hole in this wall and a mirror arrangement. Another small opening in this wall permitted *E* to insert a forceps into the interior of the apparatus to manipulate the food in the feeding treatments.

During a 4-day period of adaptation to the feeding apparatus, each *S* was put into the box twice a day for 2 min., with the forceps inserted. No food was held in the forceps at this time. Upon completion of the 2-min. adaptation period, *S* was placed in its living cage where it found a 4-gm. pellet of Purina laboratory chow, and was allowed to eat as much of the pellet as was possible in ½ hr.

Upon completion of the adaptation period, the frustrative-feeding procedures began.

Each of the two groups, L and S, was randomly divided into two subgroups, making four groups with 10 *S*s in each. One of the two subgroups from Group S and one from Group L were trained to be inactive in response to the frustrative-feeding situation (Cond. I); the other subgroups from Groups S and L were trained to be active (Cond. A).

In Cond. I, a single *S* was put into the feeding apparatus. A 1-gm. piece of chow, held in a pair of forceps, was inserted into the feeding box and *S* was allowed to pull at the food and attempt to bite into it initially, but was not allowed to have and eat the pellet until its gross overt activity had ceased. The period of required inactivity was gradually increased until *S* remained inactive in the presence of food for at least 6 sec. In Cond. A, *S* was treated in the same manner as above, except that it was allowed to obtain the food only while struggling vigorously for it.

After each trial, *S* was returned to its home cage and found food. For the first 12 days, while *S* was on an 11½-hr. deprivation schedule, frustrative-feeding procedures were carried out twice a day after which *S* received a 4-gm. pellet for ½ hr. After this, when it was thought that *S*s could survive longer periods of deprivation, a 23-hr. feeding schedule was instigated and maintained throughout the remainder of the experiment. The frustrative-feeding procedures were then administered once a day, after which *S* found two 4-gm. pellets, and was allowed to eat for a period of 1 hr.

Stage 3.—In order to evaluate the effectiveness of procedures of Stage 2, all *S*s were run through an independent test of activity on Days 102 to 107. In particular, interest was focused on the possibility that "spontaneous activity" could be affected by a generalized form of the stimulus change (increase of illumination) associated with active or inactive behavior established in the prior stage of training. However, the major stimulus to activity or inactivity, temporarily inaccessible food, was not involved in this test.

Groups of 10 *S*s were brought in carrying cages into a windowless interior room which contained 10 individual stabilimeter-type activity cages of a design adapted from Campbell and Sheffield (1953). This apparatus has recently been described in detail by Amsel and Work (1961). Each cage sat in what was essentially an open-type box of white Celotex in order to provide a nearly homogeneous visual field and to minimize intercage visual and auditory stimuli. The range of temperature in the course of this portion of the experiment was 78° to 81°. The cages were arranged in a circular array below two ceiling lights. The first light, illuminated during a 1-hr. adaptation period was 7½ w.; the second light which produced the sudden stimulus change was a 200-w., semisilvered bulb.

The *S*s were placed individually in the 10 activity cages for a 1-hr. adaptation period under the dim (7½ w.) illumination. After this period, activity was recorded for 5 min. under dim illumination in order to obtain the prestimulus-change, basal activity levels. The bright (200 w.) overhead light was then switched on, and activity was recorded in 1-min. intervals for 5 min. The 1-min. intervals were separated by 15-sec. intervals during which time the counters were not in operation, which afforded an opportunity for *E* to record the activity for the previous minute.

At the termination of the activity test each day, *S*s were removed from the activity cages, placed in the carrying cages, and brought into a room containing the double-runway apparatus (described in detail later). Preliminary adjustment to this apparatus was carried on during this 6-day period. Each *S* was allowed to explore the double runway singly for 5 min. with all doors down, and the photoelectric system operating. They were then returned to their home cages.

All 40 *S*s were tested in this manner. In view of the fact that only 10 *S*s could be tested at one time, a latin square design was used in order to control for any possible effects the time of testing in the activity cages may have had. Since this test was made in the first 4 successive days of this 6-day period each squad of 10 *S*s was tested in all of the four necessary periods which extended from 6 hr. to 1 hr. before the regular feeding treatments.

Stage 4.—When *S*s were approximately 107 days old, the frustrative-feeding treatments were discontinued and testing for magnitude of the FE was begun. Each of the four groups was randomly divided into two further subgroups. One of these groups found the food cup empty on nonreward trials; the other found the food cup covered with a screen covering which allowed *S*s to see and smell food in the cup but not to eat it. This variable was labeled FN (food not present) and FP (food present but inaccessible).

The double-runway apparatus consisted of an L shaped double runway (See Fig. 1 for

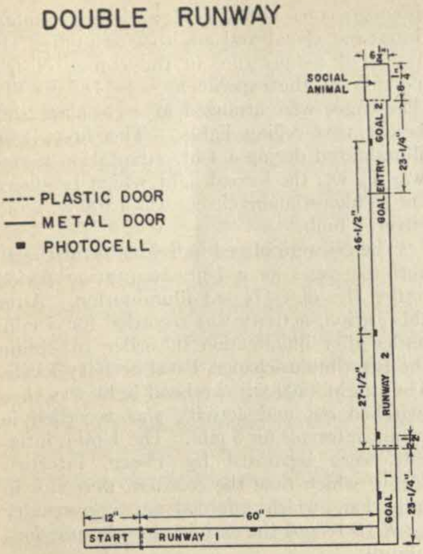


FIG. 1. Floor plan and inside dimensions of double runway.

dimensions). The first section was composed of a start box (SB), Runway 1 (R_1), and Goal Box 1 (G_1); the second section began with G_1 which served as the start box for the second runway, Runway 2 (R_2), and Goal Box 2 (G_2). It was necessary to divide G_2 into two compartments, one of which served as a temporary waiting box for a "social" *S* while the other served as a common restraining box for the social and experimental *S* when *S* entered G_2. A small, black, 1 × 1½ in. food cup, 1½ in. from the floor and approximately 2 in. from the start door of G_1, was present on all trials. A screen cover was made to slip on and off the food cup. The walls of the SB, R_1, and G_1 were black and the flooring of these sections was black rubber mat. The walls of R_2 were black and white alternating vertical stripes, and the floor was ⅛-in. brass bars spaced ½ in. apart. Box G_2 was white. The top of the apparatus was $\frac{3}{16}$-in. Plexiglas, hinged to allow access to the runways. The Plexiglas over the first section of the apparatus was overlaid with red celluloid sheeting to decrease further the level of illumination in the R_1 relative to R_2. Illumination was provided by two, 40-w. fluorescent lamps suspended in series approximately 22 in. above R_2.

The procedure followed for the acquisition of the running responses under continuous reward was essentially that of Amsel and Roussel (1952). For 18 days, each *S* was given three trials per day in the apparatus under 23 hr. hunger. During the continuous reward trials food was present at all times in G_1, and a "social" *S* was present in G_2. The details of procedure on an individual trial are as follows: (*a*) *S* was put into the SB and, after it was oriented to the door, a metal door was dropped; (*b*) 3 sec. after the door was dropped, a transparent plastic door was dropped; (*c*) *S* traversed R_1, entered G_1, and ate two small pellets (about 1 gm.); (*d*) after a 20-sec. interval, controlled automatically by an interval timer, a metal orienting door was dropped followed again by the dropping of a Plexiglas door after *S* oriented in that direction; (*e*) *S* traversed R_2 into G_2 and was allowed contact with the social *S* for 20 sec., after which *S* was removed and returned to the carrying cage. The intertrial interval was not less than 20 min.

Performance measures were taken in the second section of the apparatus by means of Standard Electric timers activated by a photoelectric system. The sequence of starting and stopping the clocks was as follows: (*a*) Twenty seconds after a photoelectric unit in the center of G_1 had been activated, a solenoid arrangement activated by an interval timer caused a metal door to drop. This was followed by the dropping of the Plexiglas start door in G_1 which started Clock 1. (*b*) When *S* moved to a point 2½ in. into R_2 Clock 1 stopped, providing a measure of latency (L), and Clock 2 started. (*c*) Passage 25 in. down R_2 stopped Clock 2, providing a measure of running time (RT), and started Clock 3. (*d*) Entrance midway into the first section of G_2, 46½ in. away, stopped Clock 3, providing a measure of goal-entry time (GT).

The 18 days of continuous reward training were followed by the test period during which each *S* continued to run three trials a day for 12 days. Of these, half were nonreward (N) trials and half were reward (R) trials. The nonreward trials differed from the reward trials in that *S* found either an empty food cup or a covered food cup containing two small pellets in G_1. Contact with the social *S* in G_2 for 20 sec. continued to be the reward for traversal of R_2. The following arrangement of daily trials was employed to control for any possible effects due to the order: NRN RNR NNR RRN NRR RNN NRN RNR NNR RRN NRR RNN. Over each 2-day block of six trials, there were three N and three R trials, one of each kind in each trial position.

RESULTS

Test of Activity to the Generalized Activity Cue

All four groups showed activity increases to the change in illumination on each test day except for a single point, the third day for Group S-I (Fig. 2). The I groups, for both small and large litters, were generally less active than the A groups. Increases in activity seem greater in the L groups than in the S groups. The prestimulus-change levels of activity are about the same for both the S vs. L and the A vs. I variables. Averaging over all 4 days, the stimulus-change activity levels were: S-A, 10.9; S-L, 8.0; L-A, 12.7; and L-I, 9.7. The prestimulus-change values in the same order were 6.3, 4.6, 5.3, and 4.5. In an analysis of variance of the prestimulus-change measures, none of the main effects was significant, *F* for between groups being less than unity. In the stimulus-change analysis the between-groups *F* was significant beyond the .01 level, but neither of the main effects (A vs. I, L vs. S) was significant at the .05 level. The interaction *F* was 4.68, significant beyond the .01 level. An analysis of covariance testing for group activity differences during the stimulus-change period adjusting for prestimulus-change differences produced an identical pattern of significance relationships. In view of the extremely large variability obtained between *S*s and the relatively small number of *S*s in each group, one might hesitantly conclude that treatments established in Stage 2 of this experiment had some effect upon the reaction to a stimulus change (illumination change) in Stage 3.

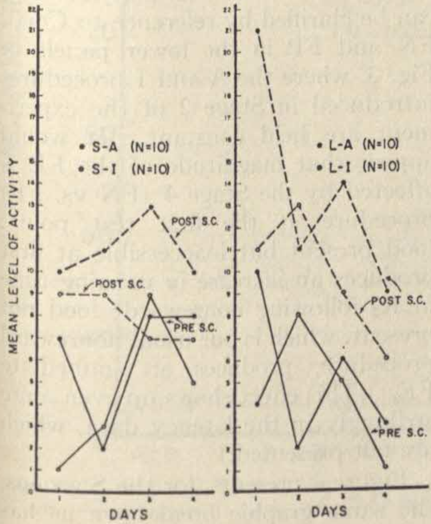


FIG. 2. Activity in stabilimeter test situation before and after stimulus change.

Test of the FE in the Double Runway

Preliminary to an analysis of variance to test differences in magnitude of the FE, *t* tests for related measures were performed to demonstrate that there was a significant FE in each of the eight groups. A median running time score over 2-day blocks of trials for R and for N trials was found for each *S* over the test period. The means of the median running times obtained on the last 6 days of testing for reward and nonreward trials were used in this analysis. In every case the difference was in the direction of the FE, and values of *t* ranged from 2.39 to 4.70, significant in each case beyond the .05 level in two-tailed tests.

Relation of FE Magnitude to Experimental Variables

Although this study was originally designed as a 2 × 2 × 2 factorial, several considerations prevented a very meaningful comparison of the magnitude of the FE in small- and large-litter groups. First, the result of an analysis of variance of the weights of *S*s 1 week before testing showed the small-litter *S*s to be significantly heavier than the large-litter *S*s. There was no significant difference in weight attributable to the frustrative-feeding (A vs. I) variable, nor was the interaction between litter size and the A vs. I feeding procedures significant. Secondly, during the introduction of the feeding procedures in Stage 2 the large-litter *S*s of Cond. I required a longer period of time to reach the 6-sec. inactivity criterion than did the small-litter *S*s; and the large-litter *S*s were more readily trained to vigorous struggling than were the small-litter *S*s upon presentation of food in Cond. A. Thirdly, as Fig. 3 and 4 show, R_2 running time during the testing was very much slower on both R and N trials for small-litter *S*s than for large-litter *S*s, although the FE shows up clearly for both groups, a finding which holds true for latency data and for goal-entry data as well. In view of the above, and the attendant possibility that the litter-size variable is confounded with a weight variable, separate statistical analyses were made for the large- and small-litter groups.

Two analyses of variance, one for L and one for S groups, were performed using the mean median running time on N trials in R_2, which for each *S* was a mean of six, 2-day (3-trial) medians. The justification for using N-trial scores alone as indicants of FE for intergroup comparisons was a between-groups analysis of R scores which yielded nonsignificant between-groups *F* ratios for R trials for both the L and S conditions. In order to equate the number of *S*s per cell of the analysis, an arbitrary *S* was added to each of the two groups in which an *S* died by using a table of random numbers and selecting the scores of 1 experimental *S* over the entire test period. A nonsignificant *F* ratio was found in both analyses for between-group means. However, the Trials × Groups interaction was significant beyond the .05 level in both analyses, and the between-trials *F* ratio was significant for the L groups, but not for the S groups.

The implications of the above analyses can be clarified by a graphic analysis of the data. Figure 3 shows a breakdown for the large-litter groups of mean median running time on R and N trials over the 12 days of testing. The two top panels compare Cond. A and I for all large-litter *S*s, holding constant the variable introduced in Stage 4 of the experiment (FN and FP treatments). The two bottom panels compare the FN and FP treatments for all large-litter *S*s. While the levels of performance on N and R trials are highly similar for Cond. A and I, suggesting that the procedures instigated in Stage 2 did not differentially effect the magnitude of the FE, these curves do show a reversal of the FE at the first points of Cond. A and I. This early reversal can be clarified by reference to Cond. FN and FP in the lower panels of Fig. 3, where the A and I procedures introduced in Stage 2 of the experiment are held constant. It would appear that magnitude of the FE is affected by the Stage 4 (FN vs. FP) procedure at the first test point: food present but inaccessible at first produces an increase in running time in R_2 following nonreward; food not present, which is our usual nonreward procedure, produces an immediate FE. (This effect shows up even more strikingly in the latency data, which are not presented.)

Figure 4 presents, for the S groups, the same graphic breakdown as has just been given for the L groups. Food present but inaccessible on nonreward trials again produces an apparent reversal of the FE for the first few days, while the FE shows immediately for the group which found no food on nonreward trials. This once more accounts for the first point reversal of the FE in Cond. A, the virtual absence of FE at the first point in Cond. I., and for the significant Trials × Groups interaction.

As indicated earlier, an analysis of variance of the reward running times

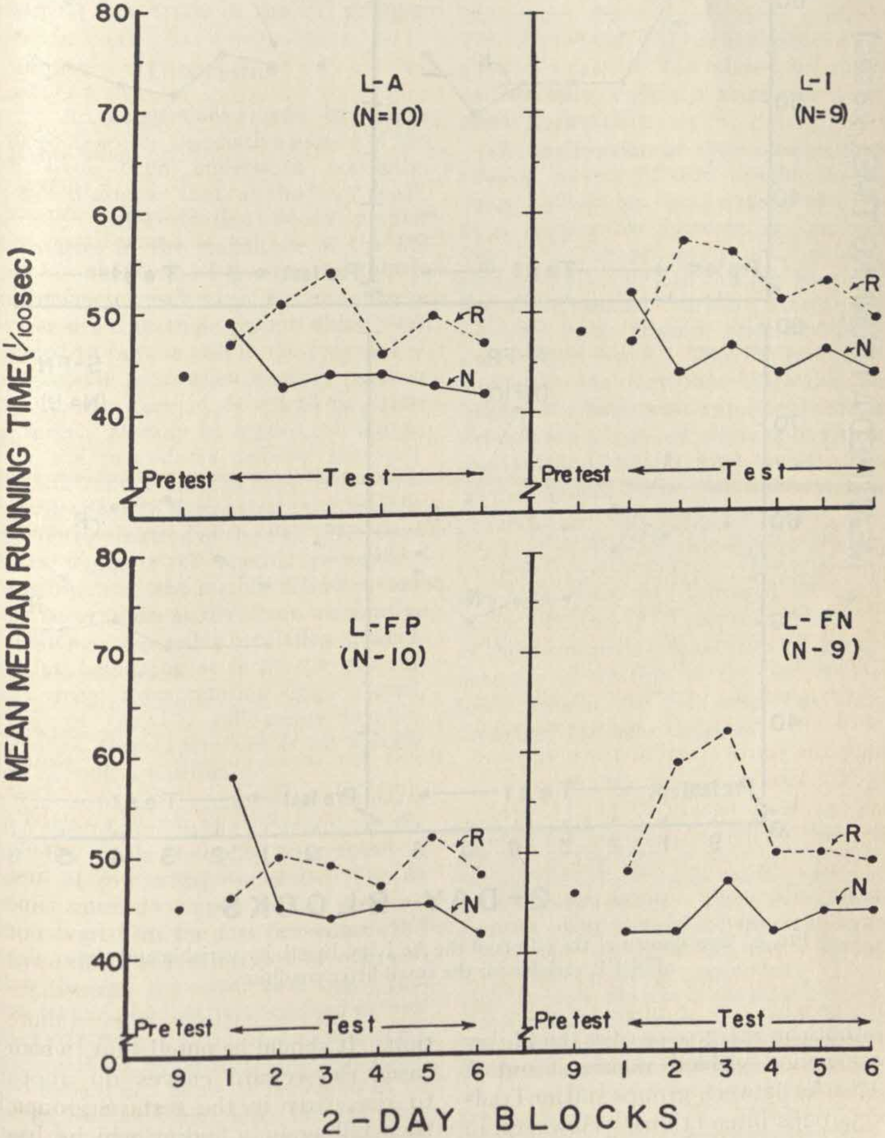


FIG. 3. Breakdown of the effects of the Activity-Inactivity variable and the FN-FP variable for the large-litter condition.

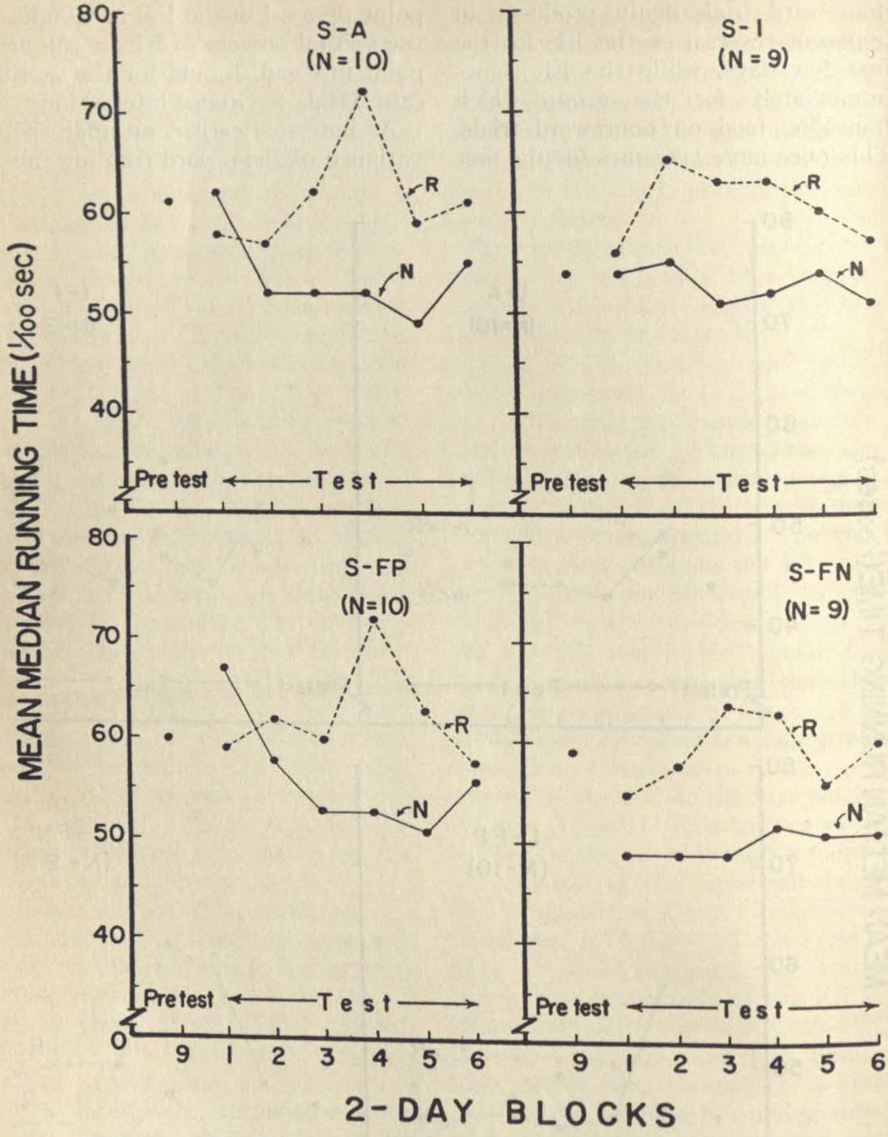


FIG. 4. Breadkown of the effects of the Activity-Inactivity variable and the FN-FP variable for the small-litter condition.

for both S and L groups over the 12-day test period yielded nonsignificant *F* ratios for between-groups and for Trials × Groups interactions. However, in both analyses, *F* was significant at the .05 level for the between-trials variation. It should be noted that in some cases the reward curves do appear to rise early in the test period and then fall again, a finding which is not characteristic of other data of this sort. A significant *F* ratio for between trials was also found in the analysis of nonreward scores for the L groups, and approached significance in the S groups. However, in the nonreward curves, the between-trial variation is more readily attributable to the decrements in running time on the first few trials in the FP groups.

DISCUSSION

If the suggestions made by Marx (1956) as to the associative nature of the FE have been understood correctly, it would appear that, at the very least, such an interpretation would predict differences in the magnitude of the FE between the A and I groups, and that it might even predict no FE or even a reversed FE in those groups which were trained to remain still in the presence of inaccessible food when hungry, particularly in the case of *S*s reared in litters of three. It may be argued, of course, that the procedures defining the variables introduced in this experiment were not adequate to demonstrate the associative determination of the FE. However, Marx suggests no specific procedures to follow, and the method followed here was to provide as divergent associative experience as possible in relation to food-getting beginning at birth. A generalized drive interpretation (e.g., Amsel, 1958) of the FE still seems tenable, although this experiment is no specific test of such a position.

The apparent reversal of the frustration effect found in the FP groups on the first few trials is readily understood in terms of competing responses. The *S*s of this condition, upon finding the food cup covered on the first few nonreward trials, vigorously attempted to bite and claw through the screen covering. This was true of the inactivity-trained as well as the activity-trained *S*s, and was often accompanied by vocalization, urination, and defection. Some *S*s remained in G_1 for as long as 4 min. after the dropping of the metal door before orienting toward R_2; and on several occasions *S*s would pass the first photocell starting the running time measure and then return to G_1. This behavior is clearly reflected in the latency data for FP groups (not presented) which show very marked increases in the group means and variability on the first few nonreward trials. As these responses disappeared, the increased running speed in R_2 after nonreward trials was clearly evident. The *S*s of Cond. FN, on the other hand, showed none of the above behavior, and the latency data of these groups are much more stable.

The implication of these observations relative to the FP-FN variable is obvious. While we have defined the FE as a measurable increase in response vigor in the second runway of the double-runway situation, and this has been our major dependent variable, there are certainly other possible effects of frustrative nonreward. The FP condition points up the fact that there can be frustrative effects which appear to reduce the criterion FE or prevent it from manifesting itself, at least temporarily. This does not mean that there is not increased *D*, but only that there are for a time stronger *H*s which this increased *D* is energizing. This produces the appearance of decreased strength of the criterion (FE) response. As soon as these competing tendencies (vigorous behavior in relation to the inaccessible food) weaken, the FE which has been "masked" becomes apparent.

SUMMARY

The major concern of this investigation was to test the effect of early experience in relation to food-getting on the magnitude of the frustration effect. Three independent variables were manipulated from the birth of 40 hooded rats until the time of testing in a double-runway apparatus. These variables were: the size of the litter in which *S* was raised; differential frustrative-feeding procedures carried out in a feeding apparatus clearly discriminable from *S*'s home environment; and no-food-present vs. food-present-but-inaccessible in the first goal box of the double-runway apparatus on nonreward trials.

A preliminary test, involving a stabilimeter apparatus, was run to evaluate the effective-

ness of the activity-inactivity training procedure.

The results of the main test in the double runway indicated the following: (*a*) Training hungry *S*s to be active or inactive in relation to food-getting does not appear to "transfer" to the double-runway situation in such a way as to influence the magnitude of the FE either in *S*s raised in large or in small litters. (*b*) During the testing period, food-not-present on nonreward trials produced an immediate and continued FE in all groups run under this condition (*c*) The presence of inaccessible food on nonreward trials appears to produce an initial increase in running times, and an apparent reversal of the FE in Runway 2. This reversal is temporary and by Day 6 of the testing period the usual FE is clearly present in all groups of this condition.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **53**, 102–119.

AMSEL, A., & HANCOCK, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126–131.

AMSEL, A., & ROUSSEL, J. S. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

AMSEL, A., & WORK, M. S. The role of learned factors in "spontaneous" activity. *J. comp. physiol. Psychol.*, 1961, **54**, 527–532.

BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951., **48**, 465–495.

CAMPBELL, B. A., & SHEFFIELD, F. D. Relation of random activity to food deprivation. *J. comp. physiol. Psychol.*, 1953, **46**, 320–322.

LAWSON, R., & MARX, M. H. Frustration: Theory and experiment. *Genet. Psychol. Monogr.*, 1958, **57**, 393–428.

MARX, M. H. Some relations between frustration and drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956.* Lincoln: Univer. Nebraska Press, 1956.

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *J. exp. Psychol.*, 1959, **57**, 130–136.

(Received November 22, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 177–182

THE PROGRESSION-REGRESSION HYPOTHESES IN PERCEPTUAL-MOTOR SKILL LEARNING[1]

ALFRED H. FUCHS[2]
Ohio State University

In a recent experiment Garvey and Mitnick (1957) demonstrated that perceptual-motor skill development may be described analytically by treating human information processing as analogous to that of a servosystem. Their analysis of lead-lag tracking errors suggested that the human operator in a continuous closed-loop control system behaves progressively, with training, as a more complicated servomechanism. The course of this progression appeared to be systematic: initially, responses to displayed system error were based on the *velocity* aspects of that input signal (i.e., its rate of change), while later in training the more subtle error acceleration and even Δ acceleration (jerk) aspects determined the control responses. This systematic development has been termed the *progression hypothesis* by Fitts, Bahrick, Noble, and Briggs (1961). This hypothesis holds that in perceptual-motor skill development *S* learns to respond to higher derivatives of the error signal's amplitude (e.g., velocity and acceleration characteristics) and to *weight* these selectively in achieving system stability and error minimization. This extension of the Garvey and Mitnick (1957) results follows from derivations of transfer functions for the human operator. Such transfer functions are descriptive of the transformations and translations imposed by the human operator on input information (system error) in generating his output (force applied to a control device). Such research, deriving human transfer functions through spectral and/or correlational analysis of tracking records, has indicated that error amplitude is more heavily weighted than error velocity in a simple positional control system (Russell, 1951, Walston & Warren, 1953), while in a more complex control system (velocity or first-order control dynamics) error velocity receives greater weight (Tustin, 1947). This implies that the initial level of control exercised by the human is the minimum necessary level to achieve system stability and that this minimum level is determined by the complexity of the system.

Further evidence for the validity of the progression hypothesis derives from research, carried out at the Goodyear Laboratories (Goodyear Aircraft Corporation, 1955), which demonstrated that two pilots of differing experiential background weighted velocity and acceleration characteristics of an error signal more heavily than error amplitude; however, only the more highly experienced *S* weighted the acceleration characteristic more heavily than error amplitude. Perhaps the most significant contribution of the Goodyear research was the direct technique employed to "fit" the human transfer function: a differential equation plus equation expressions for "neuromuscular" lag and for such nonlinearities as human thresholds and "dither" were placed in an analog computer. The *E* manipulated the human analog constants until the analog output matched the human output. The results of this matching were remarkable in that when the analog was substituted for the man in the control system, *S* (who continued to "control" unaware that he had been replaced) could not detect the substitution for periods of 20 to 30 sec. of tracking!

---

[1] This report is based on a dissertation submitted to the Graduate School of the Ohio State University in partial fulfillment of the requirements for the PhD degree. The author is indebted to George E. Briggs for his guidance throughout the course of the study, which was supported in part by Contract No. AF 30(602)-2107 between the United States Air Force, Rome Air Development Center, and the Antenna Laboratory, Department of Electrical Engineering of the Ohio State University. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] Now at Electric Boat Division, General Dynamics Corporation, Groton, Connecticut.

A corollary of the progression hypothesis is the regression hypothesis (Fitts et al., 1961) which predicts that if following skill development *S* experiences task-induced stress, weight assigned to the higher derivatives of displayed error will decrease and weight assigned to error amplitude will increase, i.e., *S* will "regress" or return to a less complex level in controlling the system. This hypothesis was suggested by the data of Russell (1951) who showed that as complexity of the input increased, relative weight assigned to error amplitude also increased. This hypothesis has received attention by Miquelon (1958) who utilized the Garvey and Mitnick (1957) technique of analyzing lead-lag tracking error scores as a means of testing the regression hypothesis. While his results were interpreted as supporting the hypothesis, they were not unequivocal. An adequate test of both the progression and regression hypotheses requires the measurement of the weights assigned to the stimulus (displayed error) characteristics and the empirical derivation of their changes during periods of training and of task-induced stress. This paper is a report of such a test.

METHOD

Apparatus.—The experimental technique was similar to that at the Goodyear Laboratories (Goodyear Aircraft Corporation, 1955) in that an analog was utilized to approximate the information-processing characteristics of the human *S* in a continuous control task. The analog was constructed to operate according to a simplified version of the human transfer function utilized in the Goodyear research (McRuer & Krendel, 1957).

The analog consisted of a differential equation of the form:

$$\theta_c + \dot{\theta}_c + \ddot{\theta}_c = (K_1\xi + K_2\xi + K_3\xi)e^{-.2p}$$

where θ_c is the human output, ξ is displayed tracking error, $e^{-.2p}$ is an exponential lag corresponding to human reaction time in continuous control tasks, and K_1, K_2, and K_3 are weighting constants for positional, velocity, and acceleration aspects of the error signal. This equation was set up on a Pace analog computer (Model 16-31R) and the servounits of a Reeves analog computer (Model C-10) hunted optimum settings for K_1, K_2, and K_3 during the tracking trials. Thus, the technique may be described as an on-line fitting of the above equation to human performance. Compared to off-line techniques previously used, the Goodyear approach is a remarkably efficient means of deriving transfer functions. A flow diagram of the system is provided in Fig. 1. The signal E^2 provided an input to the servounits and thus a servo continued to hunt for an optimum K_1, K_2, or K_3 until E^2 was reduced to zero. Since E^2 always existed at some finite level, the servo never stopped but would oscillate around the optimum value which was recorded as the value of a weighting constant for a given trial. During a particular trial only one weighting constant was estimated, the values of the other two being set at the

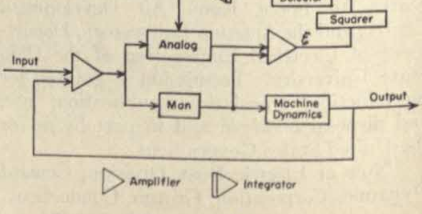


FIG. 1. The experimental system.

average of their determined values from previous trials.

Subjects and procedure.—Five undergraduate male volunteers served as *S*s and all were paid for service in the experiment. The *S*'s primary task was that of compensatory tracking; the display was an oscilloscope, the control was a spring-centered control column, and the machine dynamics through which *S*s tracked consisted of two analog integrations. A random signal provided the input, and the difference between this signal and *S*'s control movements comprised the error signal shown on the display. The *S*s were instructed to keep the cursor on target; their ultimate goal was to achieve and hold zero error.

Throughout the training period (20 days) *S* received nine trials per day; each trial had a duration of 120 sec., with an intertrial interval of 60 sec. Knowledge of their performance was accessible to all *S*s throughout the 25-day duration of the experiment; it was hoped that this would facilitate and maintain their motivational level.

After the training period, the final 5 days of the experiment constituted a transfer condition in which *S* was required to perform a secondary task simultaneously with the above-mentioned primary task; this condition is defined as "task-induced stress" (Lazarus, Deese, & Osler, 1952). The secondary task was a simple positional tracking system (Gain & Fitts, 1959) in which *S* attempted to track a .06-cpm input on a compensatory display. Instructions to *S* at the start of the transfer trials emphasized the necessity of doing as well as possible on *both* tasks. The secondary task was controlled by *S* with his left hand while tracking in the primary task with his right hand.

Performance measures utilized were (*a*) system error produced in the primary task by *S* (difference between the input and the system output), (*b*) *S*'s error on the second task, and (*c*) the gain settings used by the analog for the position, velocity, and acceleration constants of the transfer function which enabled it to match the output of *S*. The *S*'s error was not scored for the first 10 sec. of the 120-sec. trial in order to eliminate initial transients from total trial error and thus achieve a more reliable measure of *S*'s performance. The first 5 sec. of performance error on the secondary task were not scored for the same reason. Error scores during training were summarized in 2-day blocks (18 trials); the gain settings were treated similarly (6 trials per gain setting). The transfer data were summarized over single days; 9 trials for error scores, and 3 trials for each of the gain settings per day.

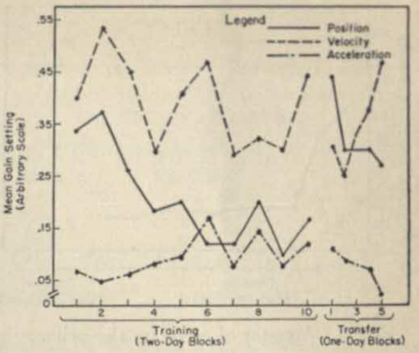


FIG. 2. Mean weights (analog gain settings) assigned to stimulus components during training and transfer.

RESULTS

Training.—The weights assigned to the stimulus components throughout the experiment are summarized in Fig. 2. The trends during training are evident: position gain, representing the weight assigned by *S*s to error amplitude, decreased during training while acceleration gain increased; the weight assigned to the velocity component of the stimulus fluctuated around a line having a generally downward trend. Analysis of these data (Table 1) reveals that these scores not only differed with respect to one another, but they also changed differentially as a function of training, as seen in the significant ($P < .01$) Measures and Measures × Days effects. The significant ($P < .01$) Sub-

TABLE 1

ANALYSIS OF VARIANCE OF THE TRAINING DATA

Source	*df*	*MS*	*F*
Measures (M)	2	11,112.495	104.58**
Blocks	9	368.332	2.14*
*S*s	4	188.650	2.41
M×*S*s	8	106.255	1.36
M×Days	18	241.782	3.08**
Days×*S*s	36	171.983	2.19**
Days×*S*s×M	72	78.398	

* $.05 < P < .10$.
** $P < .01$.

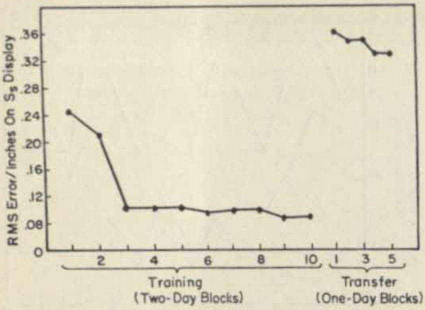


FIG. 3. Amount of error on the primary tracking task.

jects × Days interaction indicates, of course, different rates of change for these *S*s on each of the measures employed.

The usual indicator of tracking performance, the RMS error, is summarized in Fig. 3; it is readily apparent that *S*s increased their proficiency of performance on the primary tracking task. The correlation (multiple *r*) between the weights given the stimulus characteristics and RMS error was .864 ($P < .05$). The only significant single correlation was between position gain and RMS error; it was found to be .858 ($P < .01$), an interesting finding since it corroborates the prediction of the progression hypothesis that perceptual-motor skill development consists of the reduction of the weight assigned to the position component of the stimulus, with a resultant *relative* increase in the weights assigned to the velocity and acceleration characteristics.

TABLE 2

ANALYSIS OF VARIANCE OF THE CHANGES IN STIMULUS COMPONENT WEIGHTS FROM TRAINING TO TRANSFER

Source	*df*	*MS*	*F*
Measures (M)	2	1443.97	11.16**
*S*s	4	440.84	3.34
Training-Transfer (TT)	1	17.14	.28
M × *S*s	8	129.34	2.13
M × TT	2	759.57	12.51**
*S*s × TT	4	131.81	2.17
M × *S*s × TT	8	60.73	

** $P < .01$.

TABLE 3

ANALYSIS OF VARIANCE OF THE CHANGES IN STIMULUS COMPONENT WEIGHTS DURING TRANSFER

Source	*df*	*MS*	*F*
Measures	2	5978.42	21.63**
*S*s	4	392.04	2.31
Days (D)	4	103.18	.90
M × *S*s	8	204.46	1.20
M × D	8	276.34	1.63
*S*s × D	16	114.06	.67
*S*s × D × M	32	169.91	

** $P < .01$.

The weights (gains) accorded velocity and acceleration characteristics of the input were expressed as ratios of position gain in order to assess relative change during training; the asymptote of these plotted changes revealed a position/velocity ratio of .34 and a position/acceleration ratio of 1.02.

Transfer.—The shift in assigned weights from the training period to the transfer condition was analyzed (Table 2) and found to be significant ($P < .01$ for the Measures × Training-Transfer interaction). Despite the obvious relative shift in velocity and position weights *during* transfer, the Measures × Days interaction failed to attain statistical significance (Table 3); however, the over-all difference among measures was significant ($P < .01$). That *S*s performed, and did not

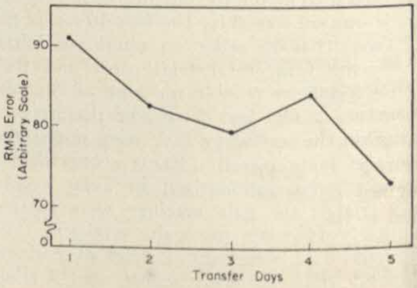


FIG. 4. Amount of error on the secondary tracking task.

ignore, the secondary tracking task is seen in Fig. 4; *S*s improved on both tasks during transfer, and thus the transfer condition may indeed be considered to have involved task-induced stress.

DISCUSSION

The training results conform to the predictions made from the progression hypothesis; at the outset of training the amplitude and velocity components of displayed error are given almost equal weight in the determination of the control responses. The decrease in position weighting, relative to velocity, indicates an early stage in the progression to a higher order of control, one in which the velocity characteristics of displayed error play an increasingly greater role in determining the response relative to the amplitude of displayed error. The lowest level of control, one in which error amplitude is the largest single determinant of the response, is not evident in these results, since the adoption of such a basis for control responses would undoubtedly have led to instability in the system. The validity of this assertion may be seen in the transfer data; on those days wherein error amplitude (position) is weighted relatively and absolutely greater than velocity, the RMS error score is higher than at any other time. The size of this error is even somewhat greater than Fig. 3 indicates, since on many of the initial transfer trials, system error surpassed the limits of the scoring mechanism; on those trials the maximum possible score was arbitrarily assigned, thus artificially limiting the size of error indicated in Fig. 3.

The relative increase in weight accorded error acceleration is even more dramatic, in that early in training it is negligible, while at the conclusion of training it approximates the weight assigned to error amplitude. That it failed to become *greater* than error amplitude is probably explicable on the basis that equivalence of these represents an optimal ratio for stability of this particular system.

TABLE 4

COMPARISON OF RATIOS REPORTED BY THE GOODYEAR LABORATORIES AND THE PRESENT RESEARCH

Pilot	Pos/Vel	Pos/Acc
VO	.71	7.14
JB	.03	.50
PR	.34	1.02

Although direct comparability with previously determined values of the relative weights for error velocity and acceleration is somewhat limited since the means of deriving these weights and the systems in which they were utilized vary considerably, it is interesting to note that the ratios of the weights reported in the Goodyear (Goodyear Aircraft Corporation, 1955) study show some similarity with the values found here. Examination of Table 4 indicates that *S*s of the present study (PR) fall between the values for the two pilots (VO represents an *S* with experience in flying light aircraft, while JB represents an *S* with considerable jet flight experience), indicating a level of control somewhat beyond the velocity order of pilot VO and approaching the acceleration level of pilot JB.

The transfer data reported in Fig. 2 substantiate the regression hypothesis; error amplitude receives greater weight during the initial days of transfer, after which the response is more largely determined by the velocity characteristic of displayed error. The decrease in the weight accorded the acceleration characteristic is probably to be accounted for by the stress condition itself. The small amount of available time to concentrate on the primary task prevented the utilization of this most subtle aspect of the display; the excessively high RMS error scores during transfer (Fig. 3) would seem to attest to this interpretation.

The direct support offered for these hypotheses also constituted indirect support for the assertion of Birmingham and Taylor (1954), Garvey and Mitnick (1957), and Fitts et al. (1961), that analogies between human performance and the performance of servosystems can provide testable hypotheses in the growth of theory for perceptual-motor skill development. The conception of such development in terms of electronic functioning seems, in the light of the results of this study, a clearly valuable contribution in that hypothetical internal characteristics of the human controller are specified and open to empirical test. The specific technique necessary to test such hypotheses has shown itself to be not only feasible but remarkably efficient; it offers rapid approximations of subtle characteristics of human responding more sensitive than the usual measures of tracking performance. That such scores *are* more sensitive is apparent in the comparison of Fig. 2 and 3; weights accorded stimulus components continue to change beyond the point at which RMS error becomes asymptotic.

The conclusions to be drawn from this study seem apparent: perceptual-motor skill performance rests upon the development with practice of the performer's ability to perceive and weight more subtle stimulus cues in the determination of more adequate error-minimizing responses. The technique utilized in this investigation holds out the promise that an adequate theory of perceptual-motor skill learning may be built upon the empirical generalizations tested herein.

SUMMARY

Two hypotheses regarding the development of perceptual-motor skill were tested. The first of these, the progression hypothesis, predicts that learning a perceptual-motor skill entails weighting more subtle aspects (velocity and acceleration characteristics) of the displayed error signal more heavily; the second, the regression hypothesis, predicts that under conditions of stress the relative weighting will "regress" to a heavier emphasis on more apparent stimulus characteristics (error amplitude), but that recovery from stress will evidence the properties predicted by the progression hypothesis. The technique employed to test these hypotheses was a system in which a human analog, consisting of analog computers and servomechanisms, was utilized to attempt to match the performance of *S*; the gain adjustments for the position, velocity, and acceleration characteristics (corresponding to the weights assigned by *S* in the system) were provided automatically by a servomechanism. These gain settings were treated as approximations of the weights assigned by *S*.

The results confirmed the hypotheses; the technique proved to be efficient and rapid, and it was concluded that the beginning of a theory of perceptual-motor skill learning may be found in these results.

REFERENCES

BIRMINGHAM, H. P., & TAYLOR, F. V. A human engineering approach to the design of man-operated continuous control systems. *USN Res. Lab. Rep.*, 1954, No. 4333.

FITTS, P. M., BAHRICK, H. P., NOBLE, M. E., & BRIGGS, G. E. *Skilled performance.* New York: Wiley, 1961.

GAIN, P., & FITTS, P. M. A simplified electronic tracking apparatus (SETA). *USAF WADC tech. Rep.*, 1959, No. 59–44.

GARVEY, W. D., & MITNICK, L. L. An analysis of tracking behavior in terms of lead-lag errors. *J. exp. Psychol.*, 1957, **53**, 372–378.

GOODYEAR AIRCRAFT CORPORATION. Investigation of control "feel" effects on the dynamics of a piloted aircraft system. *Goodyear Aircraft Corp. Rep.*, 1955, No. GER 6726.

LAZARUS, R., DEESE, J., & OSLER, S. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

MCRUER, D. T., & KRENDEL, E. S. Dynamic response of human operators. *USAF WADC tech. Rep.*, 1957, No. 56–524.

MIQUELON, D. An evaluation of a "regression hypothesis" in skilled motor performance. Unpublished master's thesis, Ohio State University, 1958.

RUSSELL, L. Characteristics of the human as a linear servo-element. Unpublished master's thesis, Massachusetts Institute of Technology, 1951.

TUSTIN, A. The nature of the operator's response in manual control and its implication for controller design. *J. Inst. elec. Engng.*, 1947, **94**, Part 2a, 190–207.

WALSTON, C. E., & WARREN, C. E. Analysis of the human operator in a closed-loop system. *USAF HRRC res. Bull.*, 1953, No. 53–32.

(Received November 28, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 183–190

THE COMPETENCE OF THE MODEL AND THE LEARNING OF IMITATION AND NONIMITATION

MILTON E. ROSENBAUM AND IRVING F. TUCKER
State University of Iowa

In imitative learning, not only does the learner respond to the cues provided by the behavior of a model but he has available potentially valuable cues in the form of the environmental consequences of the model's response. In their treatment of learning and imitation, Miller and Dollard (1941) consider societal conditions that may facilitate the imitative process, attending especially to the prestige of the models as an important relevant variable. According to their analysis, individuals who have learned to respond independently to environmental cues can, by their behavior, supply relevant cues for the behavior of others for whom the initial environmental cues are obscure, difficult, or not discriminable in some way. Models are regarded as having prestige when the imitation of their behavior consistently leads to a rewarding state of affairs for the imitator.

It is clear that a variety of learned and generalized properties of individuals are valuable clues to prestige, and hence their presence functions to elicit imitative behavior toward a potential model. Miller and Dollard allude to a variety of hierarchies extant in society within which differentiation in prestige is frequently made, i.e., age-grade levels, social status, intelligence, technical knowledge, etc. Underlying many of these dimensions is the availability of knowledge of the environmental competence of a potential model. Competence refers here to the degree to which an individual receives, in reaction to his behavior, evidence of correctness or, more apparently, reward or lack of punishment.

Previous experimental studies of imitation have confounded consistently in their experimental procedures the competence and imitative cues. The model presents his response which is to be matched by the potential imitator and in addition, he is rewarded for his behavior. In such a situation it is not possible to evaluate independently the contribution to the imitative process of the two cues. An exception to this procedural confounding appears only in the work of McDavid (1959) who, when employing adult models for child imitators, did not reward the adults.

By effecting a procedure wherein the imitative cue and the competence cue are independent of each other, the competence cue becomes essentially irrelevant to successful performance of *S*. In the present experiments *S* is capable of learning to imitate or nonimitate solely by attending to the model's response. Differential response in relation to cues of the model's competence may therefore be interpreted in terms of the pre-experimental experiences of *S*s.

It is likely, at least in our cultural milieu, that human organisms learn early to differentiate among others in terms of competence and, accordingly, in terms of prospects for successful imitation. Studies dealing with conformity behavior (Kelman, 1950; Mausner, 1954; Mausner & Bloch, 1957) have indicated the relevance of evidence of competence to the responses of an observer. The present study investigates experimentally the hypothesis that learning to imitate competent models will occur more readily than learning to imitate incompetent models. A similar analysis may be applied to nonimitative behavior.

In considering the learning of nonimitation, the most apparent and parsimonious hypothesis is simply the obverse of the imitative hypothesis: that an incompetent model's behavior will lead to responses other than imitating and therefore the learning of nonimitation of an incompetent model will obtain a higher level than the learning of nonimitation of a competent model. A variety of other suggestions may be made in predicting nonimitative response, i.e., the irrelevance of past experience with incompetent models, the tendency to nonimitate competent models who in the past received all the available rewards, and the potential salience of competency over incompetent response. The complexity of these possibilities does not permit specific prediction in the present situation.

EXPERIMENT I

Method

Apparatus.—The apparatus included two response boxes, two signal boards which presented all of the relevant stimuli to each of 2 *S*s, and a master control panel which was used by *E*. A response box was placed on the arm of a student chair in each of the two rooms occupied by the *S*s. It consisted of a double-throw lever switch by means of which each *S* indicated, upon a given signal, a "win" or a "lose" response.

A signal board placed on a table approximately 5 ft. from *S* was also duplicated in each of the two rooms. The signal board consisted of a plywood box 12½ in. wide and 18 in. high, with a depth of 6 in. It was divided into two vertical sections. The top one-third of each of these sections consisted of a compartment containing a 115-v., 7.5-w. frosted lamp, both of which could be turned on and off separately by *E*. The compartments were covered with milk glass: the left compartment was labeled "A" and the right, "B." The lower two-thirds of each vertical section was divided into two compartments, each of which also contained a 115-v., 7.5-w. frosted lamp. The milk glass covering the upper of these compartments on each side was labeled WIN and the lower compartments were similarly labeled, LOSE. All letters were 1 in. high black plastic. Between the upper one-third and the lower two-thirds of each vertical section there was a green and a red 1-in. jeweled lamp.

Procedure.—A simulated horse-racing situation adapted from that used by Brown, Clarke, and Stein (1958) was employed. Two *S*s participated in each session and each was informed by a sign on the wall of his individual room that he was Subject B. Each *S* was thereby led to believe that the other *S* was Subject A. The *E*, seated in a third room in front of a central control panel, read instructions into a microphone. The instructions indicated that when a signal lamp designating either A or B was illuminated, the appropriate *S* was to guess the outcome of an imaginary horse race. The *S* was to think of the race as involving his horse running against a field of horses, and his task was to guess whether or not his horse would win or lose the race by moving the response lever appropriately. A and B were to think of their horses as running at all times on different tracks in different parts of the country, and as never competing against each other. The *S*s were also informed that *E* was interested in the group product as well as individual responses. This information was presented in keeping with the finding by Kanareff and Lanzetta (1958) that imitative behavior is facilitated by instructions implying positive sanctions for matching.

The *S*'s guess would be indicated on the signal board in each of the rooms. The accuracy of his guess would then be shown by the jeweled lamps: illumination of the green lamp indicated a correct response; the red lamp, an incorrect response. A single trial included the races of both A and B. Subject A responded first on each trial. Subject A's win or lose response and the red or green lamp remained illuminated during B's portion of the trial.

The similated responses of A were presented by *E*. In every block of 10 trials half of *A*'s responses were win, and half were lose, in a random pattern. The designation of the correctness of the responses of A was made according to programed schedules which made up the experimental conditions for the differential competence of the model. For one-third of the *S*s, A was designated correct on 80% of the trials in each block of 10; for one-third A was correct on 50% of the trials; and for one-third A was correct on 20% of the trials.

The designation of the correctness of B's (*S*'s) win or lose response was dependent both on A's response and on the training condition. When imitation was to be trained, a response by *S* which matched that of A was correct. When nonimitation was to be trained, a response by *S* matching that of A was incorrect.

Subjects.—The three levels of the model's competence were presented in conjunction with each of the training procedures: imitation, and nonimitation. Seventy-two male students (12 *S*s in each of six conditions) were recruited in pairs from the elementary psychology course. The *S*s were exposed to 10 blocks of 10 trials, and following these, completed a questionnaire adapted from that employed by Kanareff and Lanzetta (1958) which inquired about various aspects of the experiment.

Results and Discussion

The mean frequencies of imitative responses per block of 10 trials for each of the conditions are presented graphically in Fig. 1. It is apparent that the training conditions were effective in producing differential acquisition of imitative responses. Training for imitation yielded consistently increasing frequencies of imitative responses while decreasing frequencies of imitative responses appear under training for nonimitation.

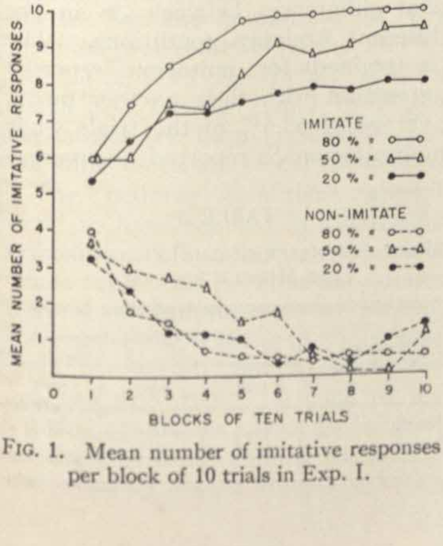


FIG. 1. Mean number of imitative responses per block of 10 trials in Exp. I.

TABLE 1

ANALYSIS OF VARIANCE OF FREQUENCY OF IMITATIVE RESPONSES IN EACH TRAINING CONDITION: EXP. I

Source	*df*	Training Conditions			
		Imitation		Nonimitation	
		MS	*F*	*MS*	*F*
Model's competence (MC)	2	87.85	3.99*	9.70	1.00
Error (B)	33	21.99		14.77	
Trials (T)	9	44.36	23.59**	33.99	16.67**
MC × T	18	2.01	1.07	2.16	1.06
Error (W)	297	1.88		2.04	

* $P < .05$.
** $P < .01$.

Summaries of analyses of variance of the data for each of the training conditions are presented in Table 1. Under training for imitation, the main effects for Model's Competence and for Trials are significant but the Model's Competence × Trials interaction is not. The significant Trials effect reflects the differential operation of the training procedures.

In order to evaluate the experimental hypothesis for learning of imitation in relation to competence of the model the critical difference technique (Lindquist, 1953) was applied to the means presented in the upper row of Table 2. Only the difference between the 80% competence mean and the 20% competence mean is significant. The mean for the 50% competence condition is intermediate. Support is offered for the hypothesis that learning to imitate a competent model proceeds more rapidly than learning to imitate an incompetent model.

The analysis of the data for training to nonimitate yields a significant

TABLE 2

MEAN FREQUENCY OF IMITATIVE RESPONSES IN EXP. I

Training Condition	Model's Competence		
	80%	50%	20%
Imitation	9.07^{a}	8.13^{ab}	7.36^{b}
Nonimitation	1.30^{c}	1.78^{c}	1.26^{c}

Note.—Means in each row marked with a letter in common do not differ significantly at the .05 level by the critical difference technique (Lindquist, 1953).

Trials effect only. This reflects the decreasing tendency to present imitative responses. The nonsignificant main effect for Model's Competence indicates failure to obtain support for the hypothesis that variations in Model's Competence lead to differential patterns in the learning of nonimitation.

Evaluation of the hypotheses was somewhat handicapped by the apparent ceiling imposed by the structure of the task on continued acceleration of learning in both training conditions. A considerable number of *S*s in each of the experimental conditions reached the maximum of either 100% imitation or 100% nonimitation early in the series of trials. Since the procedure allowed *S* in each of the training conditions to be rewarded on every occasion on which *S* responded correctly for his training condition, the frequency of ceiling level performance suggests a rather too simple task. Reduction of the frequency with which *S* can be correct for performing the response that is being trained may reduce the frequency of ceiling level performance and facilitate comparison of the effect of the Model's Competence conditions. Experiment II was conducted in accord with this suggestion.

Before proceeding to Exp. II, data derived from the postexperimental questionnaire will be considered. Ten questions were asked dealing with *S*'s expectations for performance, reports of A's and *S*'s own performance, and reports on *S*'s feelings of frustration and competitiveness. In response to the questions, *S* checked a continuous rating scale ranging from 100% to 0%. Only two of the questions resulted in statistically significant differences among the experimental conditions.

The *S* was asked to report on the model's competence in a question reading, "What percentage of his predictions did your partner predict correctly?" The means for the experimental conditions are presented in Table 3. Analysis of variance indicated a significant main effect only for the Model's Competence variable. The *S*s' reports conform in their pattern to the procedural manipulation of the model's correctness although a tendency toward centrality is apparent. The highly competent model is described as less competent than his actual performance and the reverse is reported for the least competent model.

A question asking for a report on percentage of agreement of responses of *S* and the model resulted in significant differences between *S*s in the different training conditions. The *S*s trained for imitation reported agreement with their partner on an average of 68.6% of the trials while nonimitation *S*s reported an average

TABLE 3

MEAN ESTIMATES OF THE PERCENTAGE OF THE MODEL'S ACCURACY

Training Condition	Model's Competence		
	80%	50%	20%
Imitation	71.3	48.8	37.1
Nonimitation	63.8	50.0	36.3

of 20.75% agreement ($F = 55.12$, $P < .001$). The *S*s are capable of reporting their behavior during the course of the experiment.

EXPERIMENT II

Method

The procedure was identical with that of Exp. I with two modifications. First, for both training conditions, imitation and nonimitation, the maximum frequency of designation of correctness for the prescribed instrumental response was 8 trials of each block of 10 trials. Two trials in each block were chosen at random to be designated as incorrect if the training response was made. Second, because of the possibility that acquisition of the instrumental response would be retarded by the more difficult task, the number of trials was increased from 100 to 150 trials.

The *S*s were 72 male students (12 *S*s per condition) in the elementary psychology course, and they served in pairs. Questionnaire data are not available for this experiment.

Results and Discussion

Figure 2 presents the graphical representation of the mean frequencies of imitative responses per block of 10 trials for each condition. It is apparent that the results are similar to those obtained in Exp. I but clearly altered by the imposition of the modified probability of reinforcement. In contrast to the rather smooth acquisition curves of Exp. I, the learning curves for Exp. II suggest considerable block by block variability.

The training conditions appear again to have resulted in differential acquisition of imitative responses. While the curves obtained under training to imitate tend to ascend, those reflecting training to nonimitate tend to descend. Summaries of the analyses of variance for each of the training conditions appear in Table 4.

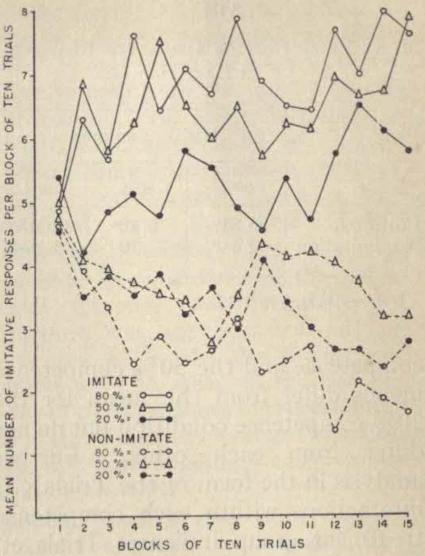


FIG. 2. Mean number of imitative responses per block of 10 trials in Exp. II.

The results for training to imitate indicate a significant Models × Trials interaction as well as significant main effects for both Models and Trials. The significant interaction term indicates that the three curves differ from each other at least on some of the trials. The critical difference technique applied to the means in the upper row of Table 5 (Lindquist, 1953) indicates that both the 80%

TABLE 4

ANALYSIS OF VARIANCE OF FREQUENCY OF IMITATIVE RESPONSES IN EACH TRAINING CONDITION IN EXP. II

Source	*df*	Training Conditions			
		Imitation		Nonimitation	
		MS	*F*	*MS*	*F*
Model's competence (MC)	2	112.03	5.93**	59.57	4.12*
Error (B)	33	18.88		14.47	
Trials	14	12.34	5.04**	11.98	5.62**
MC × T	28	3.98	1.62*	2.90	1.36
Error (W)	462	2.45		2.13	

* $P < .05$.
** $P < .01$.

TABLE 5

MEAN FREQUENCY OF IMITATIVE RESPONSES IN EXP. II

Training Condition	Model's Competence		
	80%	50%	20%
Imitation	6.84^{a}	6.40^{a}	5.31^{b}
Nonimitation	2.64^{c}	3.79^{d}	3.36^{cd}

Note.—See note to Table 2.

competence and the 50% competence means differ from the mean for the 20% competence condition but do not differ from each other. Further analysis in the form of the Trials × *S*s interactions within each competence treatment group indicates Trials effects ($df = 14/54$) significant at beyond the .01 level for both the 80% ($F = 4.65$) and the 50% ($F = 2.70$) Model's Competence conditions but a nonsignificant effect for the 20% ($F = 1.54$) condition. It appears that learning of imitation occurred under the condition in which the model's competence was high or moderate but not when the competence of the model was low.

The results for training to nonimitate indicate significant main effects both for Model's Competence and for Trials. The interaction of Model's Competence by Trials is not significant. A significant difference is obtained by the critical difference technique (lower row of Table 5) only for the comparison between the 80% and the 50% Model's Competence conditions. Trials × *S*s analysis indicates significant Trials effects for the 80% ($F = 6.17$, $P < .01$) and the 20% ($F = 1.83$, $P < .05$) Model's Competence conditions while the 50% group failed to reach a significant level ($F = 1.47$). Learning of nonimitation was obtained when the model was presented as high or low in competence but not when a moderately competent model was offered.

DISCUSSION

The findings of the present experiments tend to support those of Kanareff and Lanzetta (1958) who demonstrated that the removal of negative sanctions for imitating permits laboratory demonstration of the learning of imitation. In contrast to several previous studies (Bush & Mosteller, 1955; Schein, 1954) imitative learning was accomplished by most *S*s participating in the present studies.

The results indicate that in addition to the specific cues relevant to instrumental response acquisition, information revealing the competence of the model affects imitative learning. In order to perform at a maximum level of instrumental correctness in the present experiment, only the information indicating the response choice of the model need have been attended to by *S*s. As in previous experiments on imitative learning, designation of correctness was contingent on whether *S* matched (or did not match in the case of nonimitation) the response of the model. The correctness of the model's response was essentially irrelevant to successful matching but the results indicate that the variations in model's competence were contributory to the acquisition process. This clearly suggests that *S*'s prior social learning history generalized to the present situation.

As suggested by Miller and Dollard (1941) it is likely that early in life, discriminative learning takes place in which situations, models, drives, and other stimuli are differentiated in terms of their instrumental association with imitative behavior. Although, as with other dimensions, differentiation of models may occur for a variety of attributes, most of these connote some degree of environmental competence. Age, brightness, status, skill, are among the attributes suggested by Miller and Dollard (1941).

The effect of early experiences in the learning of nonimitation may involve several complex features. Although an apparent hypothesis is that individuals learn that incompetent models serve as useful guides to nonmatching behavior, it is also likely that in earlier learning experiences the cues provided by incompetent behavior are primarily irrelevant to successful responses. Competent models, on the other hand, in some situations may present rather reliable cues for nonmatching. This will be true especially in situations wherein the model's competence enables him to claim all the available rewards, leaving the observer only the alternative of a nonmatching response. Alternatively, competence may be consistently highlighted during early experiences so that all cues associated with competence are more salient than those related to incompetence. While Exp. I presents no differences among Model's Competence conditions in learning to nonimitate, the data of Exp. II indicate facilitation of learning of nonimitation in the 80% Model's Competence condition in comparison to the 50% condition. A trend in the direction of facilitation of acquisition of the nonimitative response appears for the 20% Model's Competence group but statistical significance is not reached. The results for this group may reflect the conflicting response tendencies based on the competing early experiences with incompetent models that were suggested above.

Kanareff and Lanzetta (1958), Lanzetta & Kanareff (1959), and Niemark and Rosenberg (1959) found that imitative response did not achieve correspondence with the probability of designation of correctness to which *S*s were exposed. Although in each of the present experiments nonvarying probabilities were employed (100% in Exp. I and 80% in Exp. II), in each case the matching of the probabilities of correctness and imitative response was quite close for several of the experimental groups. In Exp. I under training to imitate, *S*s in the 80% Model's Competence condition imitated during the last five blocks of trials at a level continuously exceeding 98%. On only six of the trials was there failure to match on the part of any *S*s in this group. All but 6 *S*s trained to nonimitate achieved levels of response of nonmatching approximating the probability of correctness. The 6 *S*s were distributed over the three Model's Competence conditions.

In Exp. II coincidence of the instrumental response and probability of correctness was approximated only by *S*s in the 80% Model's Competence conditions. The mean of the last five blocks of trials for those in this condition who were trained to imitate was 7.37, while the corresponding mean for nonimitators was 1.91.

Another finding that is in contrast to results obtained by Kanareff and Lanzetta (1958), Lanzetta & Kanareff (1959), and Niemark and Rosenberg (1959) concerns differential rapidity of learning of imitation and nonimitation. Their results indicate more rapid acquisition of imitation than nonimitation. These findings are not repeated in the present data and, in fact, a trend in the direction of more rapid acquisition of the instrumental response is indicated for the nonimitation condition in the two experiments (cf. Fig. 1 and 2). This is consistent with Bush and Mosteller's (1955) suggestion that events which increase nonimitation have a greater effect than events that increase imitation.

The present experiments differ in two important ways from previous studies in which conformity behavior or matching has been examined in relation to the competence of the model (Kelman, 1959; Mausner, 1954; Mausner & Bloch, 1957). In the present situation, the course of acquisition of the dependent response was examined in contrast to simple measurement of its current frequency of appearance. In addition, information as to the model's competence was revealed sequentially in the context of serial response rather than as a pre-experimental manipulation.

SUMMARY

The effect of information concerning the competence of a model on learning of imitative and nonimitative behavior was examined in two experiments. The *S*s predicted the outcomes of a series of fictitious horse races after exposure on each trial to the prediction and correctness of the prediction made by a simulated partner. In three conditions varying the model's competence, the partner was correct on 80%, 50%, or 20% of each block of 10 trials.

In Exp. I, half of the *S*s were trained under conditions wherein they were correct 100% of the time if their predictions matched the response of the partner. The other half of the *S*s were correct on 100% of the trials if they nonmatched the partner's prediction. The procedure allowed independent designation of correctness by specifying that the horse races predicted in each trial were run on separate race tracks. Ten blocks of 10 trials were employed. Under training to imitate, the results indicate that the greater the model's competence, the greater is the facilitation of the learning process. Under training to nonimitate, no differences appeared among the three conditions varying the model's competence. Questionnaire data indicated that *S*s were capable of accurate report of several aspects of the procedure.

Experiment II was designed to avoid a ceiling effect by allowing *S*s to be correct a maximum of 80% of the trials, and the number of trials was extended to 150. The results for training to imitate again indicated that the greater the model's competence, the greater is the facilitation of acquisition of the imitative response. Under training to nonimitate, learning was facilitated when the model's competence was high or low but was poorer when the model's adequacy was mediocre.

The results were discussed as reflecting the effect of prior social learning experiences on current social behavior.

REFERENCES

BROWN, J. S., CLARKE, F. S., & STEIN, L. A new technique for studying spatial generalization with voluntary responses. *J. exp. Psychol.*, 1958, **55**, 359–362.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

KANAREFF, V. T., & LANZETTA, J. T. The acquisition of imitative and opposition responses under two conditions of instruction-induced set. *J. exp. Psychol.*, 1958, **56**, 516–528.

KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.

LANZETTA, J. T., & KANAREFF, V. T. The effects of a monetary reward on the acquisition of an imitative response. *J. abnorm. soc. Psychol.*, 1959, **59**, 120–127.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCDAVID, J. W. Imitative behavior in preschool children. *Psychol. Monogr.*, 1959, **73**(16, Whole No. 486).

MAUSNER, B. The effect of prior reinforcement on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 65–68.

MAUSNER, B., & BLOCH, B. L. A study of the additivity of variables affecting social interaction. *J. abnorm. soc. Psychol.*, 1957, **54**, 250–256.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

NEIMARK, E. D., & ROSENBERG, S. The effect of "social" discriminative cues on probability learning. *J. exp. Psychol.*, 1959, **58**, 302–311.

SCHEIN, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.

(Received November 28, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 191–197

EFFECTS OF MENTAL SET UPON VESTIBULAR NYSTAGMUS[1]

WILLIAM E. COLLINS

Civil Aeromedical Research Institute, Federal Aviation Administration, Oklahoma City

During a course of repeated rotational stimulation, vestibular nystagmus has been reported to undergo "habituation," i.e., the oculomotor reaction and, in man, the subjective sensation of rotation, decline (Aschan, 1954; DeWit, 1953; Dodge, 1923; Griffith, 1920; Guedry & Ceran, 1959). Wendt (1951), however, holds that such declines occur only when visual stimulation is permitted or when, in total darkness, *S* does not maintain an "environment-directed orientation." He says ". . . (in darkness) loss of nystagmus occurs when the subject fails to remain alert to his external surroundings. This habituation is preventable by keeping the subject attentive to the external surrounds and avoiding inward-directed revery states" (p. 1214).

To evaluate such a mental set variable, a study was conducted (Collins, Crampton, & Posner, 1961) to determine the effect of instructions, which would influence *S*'s mental activity, on the nystagmic response elicited by angular acceleration. Significantly less nystagmus was produced during stimulation when *S*s were requested to assume reverie states than when they were kept alert throughout a trial either by signaling subjective events or by doing prolonged arithmetic problems, although there were certain inconsistencies in the results. The design of this first study was such as to place considerable demands upon *S*s, i.e., (*a*) they were required to shift rapidly from mental activity to reverie states in successive trials, and (*b*) the period between instructions to relax and daydream and the start of the rotational stimulus was quite brief (30 sec.). It is likely that these factors account for the inconsistencies noted.

The present study was undertaken to clarify the effects of various types of mental states on vestibular nystagmus by establishing favorable conditions for obtaining specific sets. Tasks and instructions were designed to produce four types of subjective conditions: (*a*) mentally alert, attending to the rotational stimulus, and signaling subjective events; (*b*) mentally alert and attending to an environmental stimulus other than that of rotation; (*c*) mentally alert and ignoring the environment; (*d*) in a state of reverie and ignoring the environment.

METHOD

The stimulator, situated in a light-proof room, consisted of a turntable on which *S* was seated with his head in the center of rotation. The rotary apparatus is described in detail elsewhere (Guedry & Kalter, 1956) and was the same as that used in the previous report (Collins et al., 1961). The modifications introduced here were the employment of a buzzer and a bell located directly behind *S*, a Hunter timer used by *E* to control the duration of the buzzer, and a telegraph key manipulated by *S* and connected to both the buzzer and the recording device. A second telegraph key, used to signal subjective events, was also connected to the recorder.

---

[1] The author is indebted to F. E. Guedry, Jr., G. H. Crampton, and L. S. Caldwell for their many helpful suggestions, and to Kenneth J. Gall for his excellent technical assistance.

This study was conducted at the United States Army Medical Research Laboratory, Fort Knox, Kentucky.

Two criteria were used in the selection of *S*s: (*a*) no history of labyrinthine disorder; (*b*) no previous experience in vestibular experimentation. Three electrodes were employed with each *S*. Two of these provided measurement of horizontal eye movements and were taped on the temporal side of each eye, slightly to the rear of the junction of the temporal and zygomatic bones. The third electrode was taped behind the left ear on the mastoid process.

The amplifying and recording device was an Offner Type T electroencephalograph. Three channels were used. One of these recorded the horizontal component of vestibular nystagmus, another denoted the periods of acceleration and deceleration, and the final one recorded signals made by *S* with either of the two telegraph keys.

Prior to each trial, 20° eye-movement calibrations were obtained from each *S* and used as conversion factors for the later data measurements. All trials were conducted in complete darkness with rotation in a clockwise direction. The *S*'s eyes were open and his head held upright by means of a biteboard.

Four conditions were employed: (*a*) Key pressing. The *S* was required to signal the onset, cessation, and each experience of 90° of rotation. This technique, described by Guedry and Lauver (1961), keeps *S* attending to the sensation of rotation. (*b*) Mental arithmetic. The *S*s were instructed to ignore the rotational environment and to do mental arithmetic problems as rapidly, as continuously, and as accurately as possible. They were given a constant divisor and a starting dividend and were required to divide each successive quotient by that divisor. Answers were recorded at the conclusion of each trial. The task was designed to provide a means of keeping *S* "inwardly directed" but mentally alert and active. (*c*) Temporal reproduction. The *S*s were instructed to attend carefully to various durations of sound stimuli (buzzer) presented at frequent intervals throughout each trial. They were to listen carefully for the start of each such stimulus and, at its conclusion, reproduce the duration of the sound by means of one of the telegraph keys. Pressing this key both activated the buzzer and marked the duration of the response on the recording paper. This condition kept *S* "environmentally directed," but with his attention oriented toward time estimation rather than the experience of rotation. (*d*) Reverie. The *S*s were asked to relax as completely as possible, but to keep their eyes open. They were told that they could daydream, but were asked not to think about anything of real concern to them, nor deliberately to pursue any line of thought.

Each *S* was tested for 5 successive days. On Day 1, four practice trials, one under each mental condition, were given. The following four daily experimental sessions each consisted of six trials with rest periods of 3–4 min. interspersed between them. The same condition was employed throughout a session for a given *S*, but the conditions differed from day to day. Further, a counterbalancing of conditions was employed with 2 *S*s assigned to each of the four established orders of presentation.

The procedure for the experimental trials consisted of the following: (*a*) 1 min. of rotation at a constant velocity of 1 rpm, (*b*) 15 sec. of acceleration at $4.5°/\text{sec}^2$, (*c*) 2 min. at a constant velocity of 12.2 rpm, (*d*) 15 sec. of deceleration at $4.5°/\text{sec}^2$, and (*e*) 2 min. at a constant velocity of 1 rpm.

The key pressing and mental arithmetic conditions were further characterized by the sounding of a bell for a period of 2 sec., 12 sec. prior to acceleration. This served to signal *S* to begin work on the arithmetic problem in the one case, and to alert him to the impending start of rotational stimulation in the key pressing situation. The bell was also sounded 12 sec. prior to deceleration in the later condition for the same reason. The sound stimuli for reproduction of duration were interspersed at definite points throughout each trial. The stimuli ranged from 0.5 to 6.5 sec. at 1-sec. intervals. Each of these values was presented twice in a given trial.

RESULTS

The recordings were analyzed for two types of data: nystagmus output (degrees of slow-phase eye movement), and duration of response. These data were scored separately for acceleration and deceleration. The duration of the vestibular reaction was measured in seconds from the first nystagmic beat to the end of the primary response. The output data were scored in centimeters with a Keuffel and Esser map measure (Model 1741) and transformed to

TABLE 1

THE AVERAGE OUTPUT AND DURATION OF NYSTAGMIC RESPONSE FOR EACH OF THE FOUR EXPERIMENTAL CONDITIONS

Ss	Nystagmus Output (Degrees)				Nystagmus Duration (Seconds)			
	Mental Arith.	Key Press	Temporal Reprod.	Reverie	Mental Arith.	Key Press	Temporal Reprod.	Reverie
	Acceleration							
1	616	684	402	384	57	49	41	43
2	505	457	320	55	39	35	32	7
3	708	561	531	364	56	43	41	33
4	361	452	189	343	40	41	24	33
5	609	509	580	510	55	49	50	47
6	930	824	750	418	51	44	46	33
7	367	225	329	44	37	31	37	17
8	408	329	218	39	31	33	27	4
Mean	563	505	415	270	46	41	37	27
SD	195	190	193	192	10	7	9	16
	Deceleration							
1	524	319	329	253	50	35	36	38
2	317	259	160	5	38	31	21	1
3	549	619	615	51	39	41	48	18
4	451	455	422	438	51	44	38	36
5	626	565	622	407	74	62	64	47
6	896	795	721	439	42	40	39	30
7	635	328	421	53	47	32	38	20
8	485	435	336	73	37	36	28	11
Mean	560	472	453	215	47	40	39	25
SD	169	180	187	191	12	10	13	15

TABLE 2

ANALYSES OF VARIANCE FOR THE NYSTAGMUS OUTPUT AND THE NYSTAGMUS DURATION DATA

Source	*df*	Error Term	Nystagmus Output		Nystagmus Duration	
			MS	*F*	*MS*	*F*
(1) Subjects (*S*)	7	4	196,765	7.91***	809	7.13***
(2) Direction (D)	1	4	2,773	0.11	0	0.00
(3) Conditions (C)	3	5	298,311	20.85***	1,159	21.19***
(4) *S* × D	7	7	24,884	4.76**	113	5.77**
(5) *S* × C	21	7	14,305	2.73*	54	2.78*
(6) D × C	3	7	6,563	1.25	14	0.73
(7) *S* × D × C	21	—	5,231	—	19	—
Total	63					

* $P < .05$.
** $P < .01$.
*** $P < .001$.

degrees of eye movement by means of the conversion factors. The measurements made represent the vertical distances from the peak of each nystagmic beat to the slow-phase baseline of that beat, accumulated for the 30-sec. period beginning with the acceleration or deceleration stimulus.

Mean values for each *S*, based on the six trials for each experimental

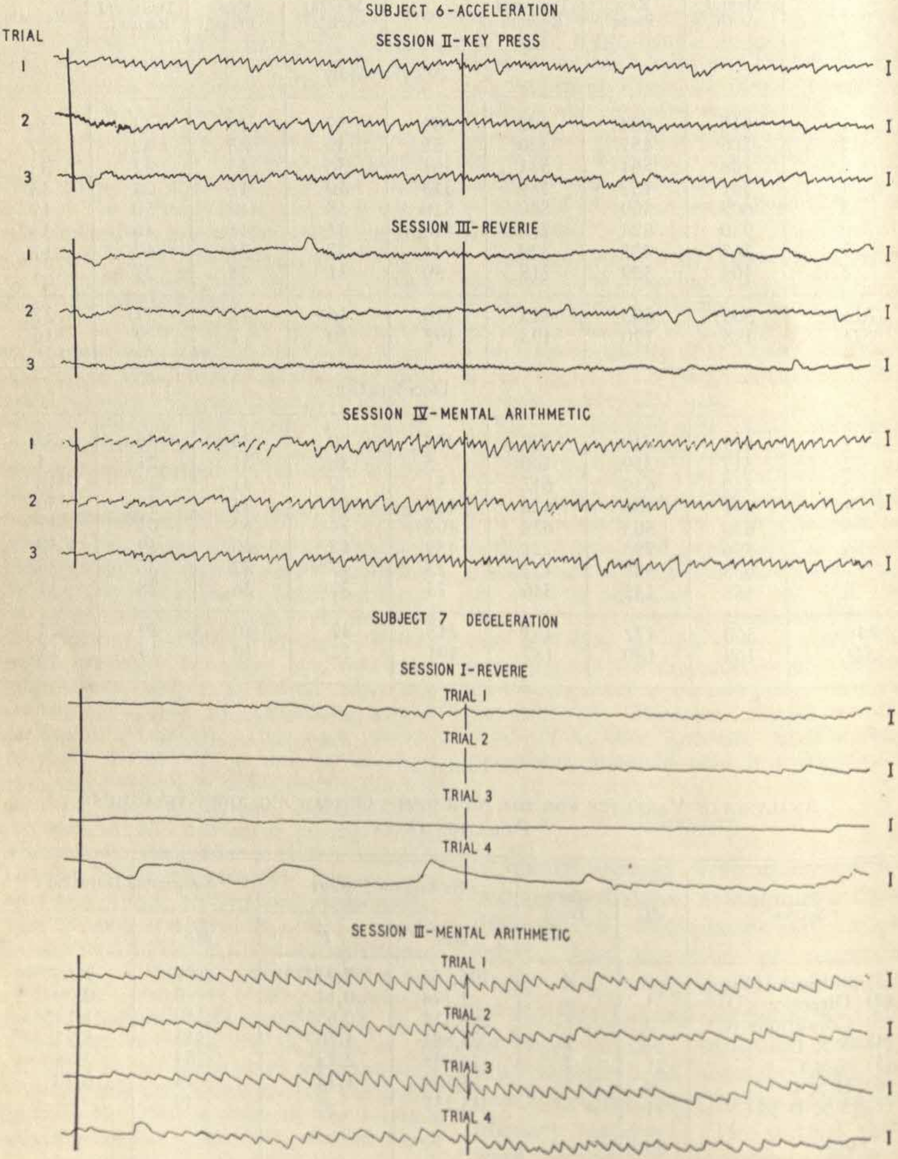


FIG. 1. The recorded nystagmus elicited by acceleration and deceleration stimuli for 2 *S*s. (Vertical bars through each record denote the 15-sec. periods of stimulation and calibration markers on the right indicate 20° eye excursions.)

condition, are presented in Table 1. These data were used in the separate analyses of variance conducted for each of the two nystagmus measures (Table 2).

The analyses and *t* tests for repeated measurements on the same *S*s yielded an identical pattern of results for both the output and the duration of nystagmus data. In neither case was the direction of stimulation (acceleration and deceleration) a significant variable. Significant *F* ratios were obtained for *S*s and for the four conditions of mental set. Since there were no directional differences, *t* tests were conducted among the conditions using pooled acceleration and deceleration scores. For both output and duration, the results of the *t* tests indicated: (*a*) significantly less nystagmus during reverie states than during any of the three attention-demanding tasks; and (*b*) significantly more nystagmus for the mental arithmetic condition than for any of the other tasks.

Some representative examples of the recorded nystagmic responses appear in Fig. 1. Both acceleration and deceleration data are presented. The marked reduction in nystagmus during reverie conditions, as well as the sharply defined, regular beats that occur during the mental arithmetic task, are readily apparent.

TABLE 3

RESULTS OF THE *t* TESTS AMONG THE EXPERIMENTAL CONDITIONS FOR THE NYSTAGMUS OUTPUT AND THE NYSTAGMUS DURATION DATA

Conditions	Mental Arith.	Key Press	Temporal Reprod.	Reverie
			Nystagmus Output	
Mental Arith.		2.74*	5.49***	5.50***
Key Press	4.27**		1.65	5.39***
Temporal Reprod.	4.89**	1.10		3.54**
Reverie	7.32***	4.26***	3.32*	
	Nystagmus Duration			

* $P < .05$.
** $P < .01$.
*** $P < .001$.

As with the main effects, a like pattern of significant interactions appeared from the two analyses of variance. In both cases, significant *F*s were obtained for *S* × Conditions and for *S* × Direction of Stimulation. These relationships indicate that the stimulus direction and the mental set conditions exert some differential effects upon different *S*s, e.g., although the overall effects of stimulus direction were not significant, some *S*s produced less nystagmus as a result of the deceleration stimulus, while others produced more, when compared to the output and duration of response to acceleration.

DISCUSSION

Nystagmus.—The data obtained by instructing *S*s to assume different states of attention point clearly to the importance of the subject's mental set in the production of vestibular nystagmus. As Wendt (1951) indicated, reverie states result in a decline of the nystagmic response, but the statement that the response will be maintained from trial to trial as long as *S* remains "environment-directed" requires qualification. If this statement is interpreted to mean that *S* must attend either to the sensations aroused by rotation or to other environmental stimuli, then the key pressing and temporal reproduction conditions, respectively, fulfill such requirements and do provide means for maintaining a high level of nystagmic output. However, mental arithmetic is at least equally effective in this regard even though the subjects are frequently unaware of the external environment, including the rotational stimulus, while working on the arithmetic problems.

What appears to make the arithmetic task so effective is that it requires a continuous and concerted state of alertness on the part of *S*. Further, at least with *S*s used here, the task assumes the nature of a challenge. The *S*s expressed concern about their accuracy and speed compared with others.

The estimates of subjective velocity made through the key pressing technique, while keeping *S*s alert, appear to have one drawback, i.e., habituation. The experience of rotation has been shown to decline with repeated stimulation (Guedry & Ceran, 1959). Under such conditions, as the sensation decays, *S*s probably tend to relax thereby reducing both the magnitude and duration of the nystagmic reaction (Guedry, Collins, & Sheffy, 1961; Guedry & Lauver, 1961).

Reproduction of sound durations ranked third in efficiency of maintaining the vestibular response. It is possible that its power could be increased. In the present study the sound level of the buzzer was quite high—such that *S*s had no difficulty in hearing the stimuli. Thus it was not necessary that *S* studiously attend to the stimulus lest he fail to hear it. It is felt that this method of maintaining alertness might be improved by using sound stimuli that are close to threshold, thereby requiring higher levels of attention on the part of *S*s.

One of the main factors, then, in the maintenance of vestibular nystagmus, would seem to be the state of alertness of *S* during stimulation. Such an explanation accounts for the differences obtained by Mowrer (1934) with "rested" and "excited" pigeons, and for other findings (Griffith, 1920) that demonstrated an enchanced vestibular response immediately after *S*s had engaged in strenuous physical exercise. In both cases, the animals and men were "keyed up" when tested. Further, the absence of a "constant reaction" noted for some *S*s in a recent study by Suzuki and Totsuka (1960) is probably attributable to fluctuations in alertness. That a "temporary reduction in response followed by a complete and lasting restoration of nystagmus" (p. 575) also occurred seems to indicate that the states of alertness of *S*s were somehow changed, perhaps by questions or by instructions.

These findings also agree with a point made previously in an investigation concerned with caloric stimulation of the labyrinth (Mahoney, Harlan, & Bickford, 1957). It was noted that when *S*s closed their eyes during a caloric nystagmus, eye movements were often completely inhibited. However, performance by *S*s of a short arithmetic task at this time "frequently elicited a brief burst of nystagmus" (p. 52).

The usual conditions under which rotational stimulation of the vestibular system is experimentally applied in total darkness tend to favor the development of a reverie attitude. However, that such states account for all of the response decline during a course of repeated rotation has not been conclusively demonstrated for man. In this regard, it should be noted that, for cat, a marked loss of vestibular nystagmus could be recovered only partially by such alerting techniques as electric shock and high intensity auditory stimulation (Crampton & Schwam, 1961). The use of d-amphetamine does not prevent this decline (Crampton, 1961).

Individual differences.—Although it is clear that reverie can strongly influence vestibular nystagmus, there are wide individual differences in the effect of this mental state. For some *S*s the nystagmus decline may be so great that the response appears to be completely absent, or the amplitude of the beats, when they occur, is so small as to preclude the possibility of scoring by the technique employed here. And yet, for 1 *S* (*S*4; see Table 1), the decline during reverie is negligible.

All *S*s reported that they had little or no difficulty in relaxing and daydreaming. However, this may mean different things, both physiologically and psychologically, to individual *S*s. Thus, it would seem that some type of objective task should be assigned to *S*s employed in vestibular research in order to assure, to the highest possible degree, uniformity of mental set.

SUMMARY

During rotational stimulation *Ss* were instructed, in different test sessions, to: (*a*) do mental arithmetic, (*b*) make estimates of subjective velocity, (*c*) reproduce durations of sound stimuli, and (*d*) assume a state of reverie. The reverie state resulted in significantly less recorded nystagmus than the other three tasks, while mental arithmetic was the most efficacious means of obtaining a vigorous response. A factor of sustained alertness seems to be an important condition for the maintenance of nystagmus.

REFERENCES

ASCHAN, G. Response to rotary stimuli in fighter pilots. *Acta oto-laryngol., Stockh.*, 1954, Suppl. No. 116.

COLLINS, W. E., CRAMPTON, G. H., & POSNER, J. B. Effects of mental activity on vestibular nystagmus and the electroencephalogram. *Nature, Lond.*, 1961, **190**, 194–195.

CRAMPTON, G. H. Habituation of vestibular nystagmus in the cat during sustained alertness produced by d-amphetamine. USA Med. Res. Lab. Rep. 1961, No. 488.

CRAMPTON, G. H., & SCHWAM, W. J. Effects of the arousal reaction on nystagmus habituation in cat. *Amer. J. Physiol.*, 1961, **200**, 29–33.

DEWIT, G. Seasickness: (Motion sickness) a labyrinthological study. *Acta oto-laryngol., Stockh.*, 1953, Suppl. No. 108.

DODGE, R. Habituation to rotation. *J. exp. Psychol.*, 1923, **6**, 1–35.

GRIFFITH, C. R. The organic effects of repeated bodily rotation. *J. exp. Psychol.*, 1920, **3**, 15–46.

GUEDRY, F. E., JR., & CERAN, S. J. Derivation of "subjective velocity" from angular-displacement estimates made during prolonged angular accelerations: Adaptation effects. USA Med. Res. Lab. Rep., 1959, No. 376.

GUEDRY, F. E., Jr., COLLINS, W. E., & SHEFFY, P. L. Perceptual and oculomotor reactions to interacting visual and vestibular stimulation. *Percept. mot. Skills*, 1961, **12**, 307–324.

GUEDRY, F. E., JR., & KALTER, H. Description of human rotation device. USA Med. Res. Lab. Rep., 1956, No. 242.

GUEDRY, F. E., JR., & LAUVER, L. S. The oculomotor and subjective aspects of the vestibular reaction during prolonged constant angular acceleration. *J. appl. Physiol.*, 1961, **16**, 215–220.

MAHONEY, J. L., HARLAN, W. L., & BICKFORD, R. G. Visual and other factors influencing caloric nystagmus in normal subjects. *Arch. oto-laryngol.*, 1957, **66**, 46–53.

MOWRER, O. H. Influence of "excitement" on the duration of post-rotational nystagmus. *Arch. oto-laryngol.*, 1934, **19**, 46–54.

SUZUKI, J., & TOTSUKA, G. Post-rotatory nystagmus: Modifications observed in experiments with repeated rotatory stimulation. *Acta oto-laryngol., Stockh.*, 1960, **51**, 570–578.

WENDT, G. R. Vestibular functions. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1191–1223.

(Received November 30, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 198–206

RESPONSE TO SIMULTANEOUS STIMULATION OF TWO SENSE MODALITIES[1]

JACK A. ADAMS AND RIDGELY W. CHAMBERS[2]

University of Illinois

Is it possible for a human *S* to do two things at once without impairment? Stated more concretely, if *S* is presented a light and a tone simultaneously, and must respond to the light with one hand and to the tone with the other, will his response measures to the light (or the tone) be the same as when each stimulus is presented separately for one-hand response in a simple, unisensory task? Basically, the foregoing statement expresses the design of the experiment reported here, and it is the one considered sufficient to examine the question. Viewed superficially, no impairment might be expected for the compound task because each stimulus source with its own sense modality and each response with its own effector system can be considered a separate, independent S-R loop. But, considered critically, the likelihood of sensory interaction for two simultaneous stimuli, central interaction, or interaction from the motor feedback of two responses, suggest reasons why performance in the complex task might be different from that of unisensory control tasks (Adams, 1961).

The present experiment was designed to be a careful laboratory test of the fundamental load-carrying capabilities of the human *S*. A discrete bisensory tracking task was used in either a unisensory version or a bisensory version where a stochastic series of simultaneous stimulus pairs was presented, one audio and the other visual, each for response by a separate hand. The scope of the investigation was extended by having some pairs perfectly predictable with learning (event certainty), and some unpredictable (event uncertainty). All events had constant duration and thus time certainty. It was hypothesized that the prediction of an event pair, both in type and time, could often free *S* of directly sensing environmental stimuli before responding and this would minimize sensory interaction as a possible source of impairment.

METHOD

Apparatus.—The research device was a version of the two-dimensional, all-visual, discrete tracking apparatus described by Adams & Xhignesse (1960), but modified to also allow two-dimensional bisensory tracking and one-dimensional tracking to either visual or auditory stimuli. In the visual dimension, *S* had three horizontally-arranged jeweled stimulus lights in front of him at eye level. The lights were red, white, and green, with white in the middle. Each light was .5 in. in diameter, and the three lights were mounted in a 2 × 3¼ in. box. Beneath each stimulus light was a small neon response feedback light that informed *S* of the position of his control, and this direct display of response feedback cues, as well as stimulus input lights, made the task one of pursuit tracking. The stimulus lights came on in a probabilistically-defined sequence and were programed on punched tape and automatically read by a motorized tape reader, Model 2, manufactured by the Commercial Controls Corporation. The *S* had to keep moving the control so the feedback light was under the lighted stimulus as much as possible.

---

[1] This research was supported by the United States Air Force under Contract No. AF 49(638)-371, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

[2] Now at the University of Arizona.

The auditory dimension was designed to have the same pursuit tracking properties as the visual dimension. There was a low, a medium, and a high pure tone (600, 800, and 1000 cps), and they came on in a probabilistically-defined order. The *S* heard them over a headset and responded with the same control as used for visual stimuli. The auditory error coding was in terms of a complex tone. When *S* had the control in the correct position he heard the pure tone but, when he was wrong, he heard a complex tone made up of two of the three fundamental frequencies. The correct stimulus was presented as a pure tone in the complex and, superimposed on it was a second tone, rapidly interrupted, whose frequency was determined by the position of the control. Thus, when *S* was in error he was provided direct feedback on the correct stimulus and the control position, and this was formally equivalent to the visual pursuit tracking task.

Response was with a 2-in. control stick mounted on a wide armrest of *S*'s chair, and *S* could have a control for one or for each hand, depending on whether he was in a unisensory or bisensory group. Either sensory dimension could be assigned for response by either hand, and angular orientation of the control could be set by *E* because the entire control assembly could be rotated and locked for any direction of control movement. The control made electrical contact by way of a small extension arm under the armrest of the chair, and the arm had a ballbearing tip that could be rolled to any one of three small contact plates set in a plastic hemispheric disk. The distance between the three contacts was negligible and this design feature, coupled with appropriate electrical circuitry, gave no dead space between control positions so *S* was always on one of three contacts and always had knowledge of control position because one of the feedback lights was always on.

Performance measurement was by six time-on-target timers, one for each of the three stimulus events in each sensory dimension. An Esterline-Angus operations recorder allowed *E* to record the time relationships between all stimulus and response events on a trial.

Experimental procedure.—Group A was a unisensory control group and responded only to auditory stimuli. Group V was the other control group and received only visual stimuli. Group AV was the bisensory experimental group where a visual and an auditory stimulus of the same duration were always presented simultaneously, the visual event for response by one hand and the auditory event for response by the other. Using a compass reference, and considering a backward-forward control movement at a right angle to *S* as 0°, a control's directional orientation was either 45° or 315°. For Group AV, both controls had either divergent or convergent lines of movement. Within each group, the direction of control movement and right or left hand was counterbalanced among *S*s. Group AV had further counterbalancing of auditory and visual stimuli with respect to control direction and hands.

Each trial was 100 sec. in duration and consisted of 50 2-sec. events or, in the case of Group AV, 50 pairs of 2-sec events. Sixty practice trials were given. There were three practice sessions, each on a different day, of 20 trials each. Intertrial rest was 40 sec., except between Trials 10 and 11 when it was 5 min. All events and rest periods were programed on punched tape and proceeded without intervention from *E*. Time on target was recorded on every trial, but the Esterline-Angus operations recorder was used only on Trials 56 and 57 to provide for a detailed analysis at a high level of practice.

Characteristics of stimulus series.—Table 1 gives the theoretical probabilities used in generating the sequence of intramodality stimulus events in both auditory and visual, as well as the conditional probabilities for intermodality event pairs showing the probability of a visual event being paired simultaneously with an auditory event. To illustrate the implications of Table 1 for the predictability of events, it will be noted that conditional probabilities for intramodality stimuli have both certain events (white light or medium tone) and uncertain events (red or green light, high or low tone). A series of visual events could look like this: red, white, red, white, green, white, red, white, green, etc. A white event followed a colored event with certainty, but a white event had an equal chance of being followed by a red or green event. The auditory events had the same patterning. When both sensory dimensions were used for Group AV, there was the added requirement for specifying the probability relationship between the auditory and visual events that were always paired and presented simultaneously, and the implication of the intermodality conditional probabilities is that the certain white event and the certain medium tone always occurred together, and the uncertain red and green lights each had

TABLE 1

THEORETICAL PROBABILITIES OF OCCURRENCE, $p(i)$, AND CONDITIONAL PROBABILITIES, $p_i(j)$ AND $p_k(l)$, FOR EVENTS OF THE VISUAL AND AUDITORY SERIES

Probability of Occurrence			Conditional Probabilities for Intermodality Event Pairs			
Visual Stimulus (Light)	Auditory Stimulus (Tone)	$p(i)$	k	l		
			$p_k(l)$	High	Medium	Low
Red	High	.25	Red	.50	0	.50
White	Medium	.50	White	0	1.0	0
Green	Low	.25	Green	.50	0	.50

Conditional Probabilities for Intramodality Events

Visual Events				Auditory Events			
i	j			i	j		
$p_i(j)$	Red	White	Green	$p_i(j)$	High	Medium	Low
Red	0	1.0	0	High	0	1.0	0
White	.50	0	.50	Medium	.50	0	.50
Green	0	1.0	0	Low	0	1.0	0

Note.—The notations i and j refer to events within a particular sense modality, and $p(i)$ and $p_i(j)$ define the organization of events for both the visual and auditory series. The notations k and l refer to two simultaneous events, one auditory and one visual, and $p_k(l)$ is the conditional probability that a kth visual event will be paired with an lth auditory event.

an equal likelihood of being paired with the uncertain high and low tones.

Both the visual and the auditory series had a redundancy of 68% (Attneave, 1959). The procedure was to generate one tape of 1000 pairs of events, and this same tape was used for each group on each of the three sessions. Group AV was given all events on the tape, but one of the stimulus sources was withdrawn for Group A or Group V. A response to a certain event was the center position of the three on the control, and responses to uncertain events were on either side of the center.

Subjects.—There were 16 *S*s in a group. The 48 *S*s were university male undergraduates who were paid for their participation. They were randomly assigned to groups.

RESULTS

Method of data summary and presentation.—The stimulus events for the three groups represent eight experimental conditions, depending on whether the modality was visual or auditory, the events certain or uncertain, and whether the event was presented alone or simultaneously with another event in the other sense modality. For each modality there were two types of uncertain events—red and green lights for the visual, and high and low tones for the audio. The measures for the two uncertain events of a modality were pooled as a common class of uncertain events because their probabilities of occurrence were the same and they posed the same response demands for *S*. The equipment provided performance measures for each of the eight conditions. Each response measure is designated by its stimulus class, and is always labeled either C or U for certain and uncertain, and A or V for auditory or visual. In addition, the measures of Group AV will have a lower case a or v to signify that a stimulus occurred simultaneously in the other sense modality. For example, a CA measure is the response of Group A to certain auditory events, and CAv is the same response measure for Group AV, but where each certain auditory event is simultaneously accompanied by a certain visual event that is being given response with the other hand.

Time on target.—Figure 1 shows the performance curves of the eight conditions. Time on target for certain events is higher than for uncertain events in all cases (Bricker, 1955). Four Groups × Trials analyses of variance (Type I mixed design—Lindquist, 1953) were performed on the final 10 trials to determine differences between single and bisensory responding. The comparisons of principal interest were between groups, and Table 2 shows that time on target for uncertain bisensory events was significantly poorer than for uncertain unisensory events, and this held for both auditory and visual stimuli. But this was not the case for certain events because Cond. CaV had a significantly *higher* performance level than CV. This superiority for bisensory certain events was true only for the visual modality because the difference in performance between Cond. CA and CAv lacked the .05 level.

Response time analysis.—Time-on-target proficiency in discrete tracking is a function of (*a*) time between the onset of the stimulus and the onset of the response, whether the response is correct or incorrect, (*b*) number of errors, i.e., movements of the control to the wrong position, and (*c*) duration of each error before it is corrected. The operations recorder on Trials 56 and 57 allowed measurement of each of these response determiners of tracking proficiency, and tabulations were made of error frequency, and the time between onset of the stimulus and the onset of the response,

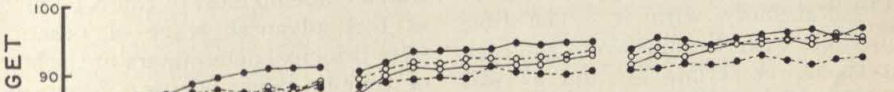


FIG. 1. Mean percentage of time on target on blocks of two trials for the eight experimental conditions. (The breaks in the curves after Trials 20 and 40 signify intersession intervals.)

TABLE 2

F RATIOS OF FOUR GROUPS × TRIALS ANALYSES OF VARIANCE OF THE TIME-ON-TARGET SCORES OF TRIALS 51–60

Source of Variance	*F* Ratio			
	CV & CaV	CA & CAv	UV & UaV	UA & UAv
Between groups	5.30*	0.60	6.28*	49.02**
Trials	1.33	2.03*	5.36**	2.58*
Groups × Trials	1.26	1.65	2.04*	1.04

* $.01 < P < .05$.
** $P < .01$.

whether the response was correct, incorrect, or correction of an error. These measures are called *response times* (RT) to distinguish them from reaction time measures where steps are taken to insure that response follows the stimulus. In a discrete tracking task a response can either follow or be anticipatory and occur before the onset of the stimulus or simultaneously with it. The time between each stimulus event and each response on Trials 56 and 57 was measured in milliseconds. The convention was followed of assigning a positive sign to RT if the response preceded the onset of the stimulus, and a negative sign if it followed the stimulus (Adams & Xhignesse, 1960; Poulton, 1952). Three types of RTs were recorded. A Correct RT was the time between onset of the stimulus and onset of the response when response was the correct one for the stimulus presented. Error RT was measured in the same way but when a response was initially wrong. Error Correction RT was the time between the onset of the stimulus and the time taken to complete the correction of a wrong response. Each *S* on Trials 56 and 57 contributed RT values for 50 certain and 50 uncertain events to the analysis. The algebraic mean in milliseconds was taken of the values of each of the three RT measures for each *S* in an experimental condition, and this gave the basic RT scores on which analyses were performed.

Table 3 gives the group algebraic mean for each of the three RT scores in each experimental condition. There were no Error or Error Correction RT scores for certain events because *S* had only two choices when a certain event appeared: (*a*) either leave the control where it was, or (*b*) move it to the correct position. All *S*s always did the latter. All mean RT scores are negative, signifying that on the average they tended to occur after the onset of the stimulus, and this is the same tendency found by Adams and Xhignesse (1960) for two-dimensional, all-visual discrete tracking. The rank order of means in Table 3 for Correct RT scores is about the same as time on target and it is inferred from this that Correct RT measures, which made up most of the RT values at this advanced stage of practice, were principal determiners of the level of time-on-target curves. Mean differences between groups for Correct RT scores were evaluated by the *t* test for independent measures, and,

TABLE 3

GROUP ALGEBRAIC MEAN RT VALUES IN MILLISECONDS FOR THE THREE TYPES OF RESPONSES MADE IN EACH EXPERIMENTAL CONDITION: POOLED TRIALS 56 AND 57

Experimental Condition	Types of Responses		
	Correct RT	Error RT	Error Correction RT
CA	− 85	—	—
CAv	− 97	—	—
CV	−175	—	—
CaV	− 98	—	—
UA	−272	−264	−520
UAv	−364	−412	−780
UV	−356	−322	−614
UaV	−357	−403	−729

as with time on target, response to visual stimuli under bisensory conditions was superior to that in the unisensory visual task. The null hypothesis test of the mean difference in Correct RT scores for Cond. CV and CaV showed CaV to be significantly faster ($.02 < P < .05$). For uncertain events Cond. UaV had a performance level significantly below that of Cond. UV when time on target was the measure, but the comparison lacked significance at the .05 level when the measure was Correct RT score. Apparently this was because *S*s under Cond. UaV had a tendency to make more errors and take longer to correct each of them, and this produced a greater separation of time-on-target means. The mean frequency of errors for *S*s under Cond. UaV was 4.9, but only 1.50 for Cond. UV ($.05 < P < .10$). Table 3 shows that *S*s under Cond. UaV had a longer mean Error Correction RT than UV, but the mean difference was not significant either ($.30 < P < .40$). Despite the lack of statistical support, it seems reasonable to believe that the combined effects of these two error tendencies served to produce mean differences in the time-on-target measure that were absent when Correct RT scores were evaluated.

The mean difference between Correct RT scores for Cond. UA and UAv, like time on target, had superior performance for UA ($P < .01$). An added reason for the superiority of UA with time on target was that *S*s under Cond. UAv tended to make more errors and took longer to correct them. The mean frequency of errors for Cond. UAv was 10.1, and was 7.2 for UA, but this difference lacked significance at the .05 level. But a *t* test of the mean difference between Error Correction RT scores of Cond. UAv and UA showed that the tendency to take longer to correct an error under Cond. UAv was statistically significant ($P < .01$).

Mean Error RT scores in Table 3 have the same general ordering as Correct RT scores and *t* tests for related measures between mean Correct RT and mean Error RT scores for each of the four uncertain conditions produced no significant differences. When *S* made an initial response to a stimulus, whether the response was right or wrong, the RT was about the same and there was no basis for the a priori possibility that wrong responses might have different characteristics than correct responses.

Ordering of response pairs.—The greater speed of response to auditory than to visual stimulus has been known for a long time (Teichner, 1954), and this was true for the present study too. For Correct RT scores, the mean difference between Cond. CA and CV, and UA and UV, favored auditory events and a *t* test of each was significant ($P < .01$). Certain audio events in the bisensory task (CAv) induced about the same mean level of Correct RT score as the unisensory certain auditory events (CA) but, interestingly, the ordinarily slower response to visual events was speeded up in the bisensory task so that the mean of Correct RT scores for Cond. CaV is about the same as that for Cond. CA and CAv. Conversely, for uncertain events, Cond. UAv had performance significantly slower ($P < .01$) than the UA unisensory control in terms of Correct RT scores, and had about the same mean Correct RT score as Cond. UV and UaV.

The common mean speed of responding for a response pair in the bisensory task, with the speed determined by the predictability of the stimulus event pair, suggested a synchronization of the two responses. To check this, the recorder charts were evaluated for the order of each pair

TABLE 4

For Group AV, Mean Percentage of the Three Possible Orders of the Two Responses to Certain and Uncertain Stimulus Events

Response Order	Type of Stimulus Event Pair	
	Certain	Uncertain
Visual before auditory	10(0)	16(2)
Auditory before visual	11(0)	10(7)
Visual and auditory simultaneous	79(0)	46(19)

Note.—Primary entry is when both responses are correct. The value in parenthesis is the percentage when either one or both of the responses of the pair are initially in error. Based on 50 certain and 50 uncertain event pairs given each *S* on pooled Trials 56 and 57.

of responses to a simultaneous pair of stimuli. One response could precede the other, or they could occur together, and the criterion of togetherness used was a difference between RTs for the two responses that was in the range ±33 msec. When the difference was larger than this, the responding was considered sequential rather than simultaneous. Table 4 gives the mean percentages of the three possible orders of response for Group AV on Trials 56 and 57. Percentages are given for certain and uncertain stimulus event pairs when both responses are correct, and when one or both are initially in error. The predominant tendency was for simultaneous response movements, with the tendency being somewhat stronger for certain events. Given this joint responding in the bisensory task, it is concluded that the response pair was geared to the faster auditory stimulus when events were certain, and to the slower visual stimulus when events were uncertain.

Anticipatory responding.—Table 3 shows that the mean Correct RT scores of less than 100 msec. for Cond. CA, CAv, and CaV, are faster than the most optimally expected reaction time values (e.g., Klemmer, 1956; 1957) and therefore it was reasonable to infer that *S*s were predicting events rather than waiting for the stimulus to occur before responding. Helson (1949) was perhaps the first to observe that learned anticipation can enhance tracking proficiency by resulting in RTs less than optimum reaction time values. Bartlett (1951) and Adams (1961) discuss anticipation as a salient feature of skilled performance, and Poulton (1952, 1957) has introduced the term *perceptual anticipation* to describe the phenomenon. Recognizing that anticipation can result in excessively premature responding, with performance poorer than if *S* actually waited for the stimulus to occur, Adams and Xhignesse (1960) discuss *beneficial anticipation* where RTs result in less off-target time than expected on the basis of optimum reaction time. They further distinguish *perfect anticipation* where there is coincidence of stimulus and response occurrence.

In this study a Beneficial Anticipation RT was defined as being equal to or less than ±133 msec., and a Perfect Anticipation RT as being equal to or less than ±33 msec., which allowed for some minor deviation from exact coincidence in deference to equipment and measurement vari-

TABLE 5

Mean Percentage of Response Time Values Obtained for Certain Events on Trials 56 and 57 That Were Beneficially Anticipatory or Perfectly Anticipatory

Experimental Condition	Beneficial Anticipation RT	Perfect Anticipation RT
CA	57%	24%
CV	33%	15%
CaV	56%	25%
CAv	57%	25%

ability. Table 5 gives the percentages of RTs for the certain events on Trials 56 and 57 that meet these two criteria. The major finding was that responses made to certain auditory events (CA and CAv), or to visual events that accompany certain auditory events (CaV), had a higher percentage of both types of anticipatory responses. To test statistically the influence of auditory events to promote a higher level of anticipation, a Beneficial Anticipation RT score was defined for an *S* as the number of Correct RTs that fell within the ±133 msec. criterion, and a Perfect Anticipation RT score as the number of Correct RTs that fell within the ±33 msec. range. It was found for the Beneficial Anticipation RT score that (*a*) Cond. CA was significantly higher than CV ($.01 < P < .05$), (*b*) the amount of anticipation for audio in the bisensory task (CAv) was not significantly different from the unisensory audio task (CA), and (*c*) Cond. CaV has a mean score that does not differ significantly from that of the bisensory audio condition, CAv. From all of this Cond. CaV would be expected to have a significantly higher Beneficial Anticipation RT score than Cond. CV, and this was the case ($.01 < P < .05$). These same tests for the Perfect Anticipation RT score gave the same results except that the mean difference between Cond. CA and CV lacked significance ($P > .05$). The inference from these comparisons is that anticipation is the mechanism that speeds the ordinarily slow response to the visual stimulus so that it occurs simultaneously with the faster response to the audio stimulus.

DISCUSSION

The question we posed in the beginning does not have a single answer. Rather than finding an answer simply in terms of impairment or no impairment for response in a bisensory task, we found that response to certain visual events in the bisensory task was actually superior to the unisensory visual control group because RTs were increased to that of the faster response made to accompanying auditory stimuli. This gain for the visual modality was because the responses to certain visual-auditory stimulus pairs were usually made together and, through the potent behavioral mechanism of learned anticipation, the ordinarily slower visual RT increased to the same speed as response to auditory stimuli. Anticipation, being a learned and internal mechanism, triggered the response from within *S* and freed it from the constraints of the S-R loop via the visual system. Somewhat unexpectedly, we uncovered a set of circumstances where performance in a complex task was better than the aggregate performances of the part tasks comprising the whole.

On the other hand, when events were uncertain there was impairment because the ordinarily faster response to auditory stimuli was slowed down to occur simultaneously with the slower response to the visual stimulus. When stimulus uncertainty existed and the event had to be sensed before response could be made, anticipation was absent and the slower response governed the speed of the two response movements made together. One other item of impairment for audio response to uncertain bisensory stimuli was a significantly longer time to correct errors than for audio response to uncertain unisensory stimuli. This error phenomenon was not found for visual stimuli. Apparently this was because auditory errors were not as simply coded, but why correction responses for less than optimally coded errors should be subject to decoding delay in a bisensory task must await further research. Whatever the reason, it is another line of evidence showing that impairment in doing two things at once occurs when bisensory stimulus events are uncertain, and is in contrast to facilitating response effects when events are certain. The relevance of stimulus predictability for doing two things at once is consistent with the study of Bahrick and Shelly (1958).

One final aspect of our data deserves comment. Why did certain unisensory auditory events have a significantly higher level of beneficial and perfect anticipation than certain unisensory visual events? We can only surmise that audio events, being more difficult than visual in the early trials, produced less "rewarding" on-target time and so anticipation emerged as a learned mechanism that allowed *S* to achieve a satisfactory measure of success. This interpretation suggests the hypothesis that one of the variables determining anticipation is difficulty level of the task, in addition to regularities in the type of events and the times of their occurrence. A study by Adams and Xhignesse (1960) is consistent with this notion because they found anticipation in a two-dimensional, all-visual discrete tracking task to occur for more difficult 1-sec. events, but not for 2-sec. events. Using voluntaristic phrasing, it is as if *S* rises to the small crisis of not succeeding well enough and invokes anticipation as a behavioral device to solve his difficulties.

SUMMARY

An experiment was performed to answer the question of whether a human *S* could do two things at once without impairment. A bisensory discrete tracking task was used where a probabilistic series of simultaneous auditory and visual stimuli were presented, each stimulus series for response with a separate hand. An auditory and a visual control group each practiced only a unisensory version of the task where response was with one hand. All events were of 2-sec. duration and had time certainty, but the type of event occurring next in the series could be either certain or uncertain.

The results revealed a net superiority of bisensory over unisensory responding when stimulus events were certain. This was because *S* in the bisensory task usually made the two response movements together, and anticipation of certain events resulted in an increase in speed of the visual response time to that of the faster audio response time. But, when events were uncertain, impairment was inferred for bisensory responding because the faster audio response time was reduced in speed and synchronized with the slower visual response time.

REFERENCES

ADAMS, J. A. Human tracking behavior. *Psychol. Bull.*, 1961, **58**, 55–79.

ADAMS, J. A., & XHIGNESSE, L. V. Some determinants of two-dimensional visual tracking behavior. *J. exp. Psychol.*, 1960, **60**, 391–403.

ATTNEAVE, F. *Applications of information theory to psychology.* New York: Holt, 1959.

BAHRICK, H. P., & SHELLY, C. Time sharing as an index of automatization. *J. exp. Psychol.*, 1958, **56**, 288–293.

BARTLETT, F. C. Anticipation in human performance. In G. Ekman, T. Husén, G. Johansson, & C. I. Sandström (Eds.), *Essays in psychology.* Uppsala: Almquist & Wiksells, 1951. Pp. 1–17.

BRICKER, P. D. Information measurement and reaction time: A review. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe: Free Press, 1955. Pp. 350–359.

HELSON, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, **62**, 473–497.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

KLEMMER, E. T. Simple reaction time as a function of time uncertainty. *J. exp. Psychol.*, 1957, **54**, 195–200.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

POULTON, E. C. Perceptual anticipation in tracking with two-pointer and one-pointer displays. *Brit. J. Psychol.*, 1952, **43**, 222–229.

POULTON, E. C. On prediction in skilled movements. *Psychol. Bull.*, 1957, **54**, 467–478.

TEICHNER, W. H. Recent studies in simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–149.

(Received December 2, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 207–213

NONREINFORCEMENT AND NEUTRALIZATION OF STIMULI[1]

DAVID LaBERGE, JAMES G. GREENO,[2] AND O. F. PETERSON

University of Minnesota

A problem common to many learning situations is the determination of the effects of nonreinforcement. It is commonly inferred that a reinforced trial strengthens the response made, although there may remain some question as to whether or not this involves also a weakening of other responses. A nonreinforced trial, on the other hand, gives rise to many conflicting interpretations, including (*a*) weakening of the response made, (*b*) strengthening of a competing response, (*c*) no effect on responses, and even (*d*) strengthening of the response made.

When applied in the two-choice prediction situation, where *S* predicts event E_1 by making response A_1 and predicts E_2 by making response A_2, the nonreinforcing event, labeled E_0, occurs when neither E_1 nor E_2 follows the response. Results of some studies (Atkinson, 1956; Neimark, 1956) designed to test theoretical interpretations of nonreinforcement in this situation have supported the identity hypothesis, namely that the average effect of a nonreinforced trial is to leave the response probabilities unchanged. Anderson and Grant (1957) have reported results which are reasonably close to predictions on the basis of the identity hypothesis, but the fact that their sharing parameter for E_0 trials apparently did not reach unity prevented a clear confirmation of the identity hypothesis. They also showed a tendency for *S*s to repeat on the next trial the response which was followed by an E_0 event.

It has been shown (Estes, 1959) that the process underlying the identity operation may be (*a*) a strengthening of the response just made through conditioning by contiguity, or (*b*) a weighted combination of the weakening of the response plus strengthening of alternative responses, or (*c*) no effect. The chief concern of this study is to investigate an additional process which could result from nonreinforcement.

Experimenters commonly observe that the introduction of nonreinforcement occasions *S* to make responses in addition to those recorded by *E*. If these responses were reinforced in some way, then presumably some stimulus elements would become connected to them, and these elements would then be regarded as neutral with respect to the A_1 and A_2 responses. In terms of statistical learning theory, therefore, a nonreinforced trial might well result in the neutralization of some of the sampled choice-point cues. If in a random sample all of the cues were neutralized, then the average relative proportion of choice cues would remain constant, and no change in the response probabilities would occur on the next trial. If only a part of the sample were neutralized, the net effect on subsequent response probability would depend upon what happened to the rest of the sampled cues.

This study attempted to investigate

---

[1] This study was supported by grants to the senior author from the National Science Foundation and from the Graduate School of the University of Minnesota.

[2] Now at Indiana University.

the neutralization of cues in the simple noncontingent prediction situation. The task of S was to predict which of two lights would follow a signal. One possible instance of the neutral response class here might be S's implicit prediction that neither light would occur following the signal, even though he was required to make one of the two overt responses associated with the lights. And presumably when neither light flashed, this might have reinforced the implicit "neither light" prediction.

There were two parts in the present investigation. The first part involved comparisons of slopes of learning curves under a reinforcement ratio of 90:10, following differential training at a reinforcement ratio of 50:50. It was expected from the deductions described in an earlier paper (LaBerge, 1959b), that groups having a higher percentage of nonreinforced trials during 50:50 training should have a greater proportion of neutral elements in the population at the start of 90:10 training, and therefore should show faster rates of learning during the 90:10 schedule.

The second part of this study involved a comparison of learning performance among groups having different numbers of nonreinforced trials inserted in a 90:10 schedule. In this case, the neutralization hypothesis implies conflicting predictions, depending upon what other assumptions are made. If the average sample contains representative proportions of choice-point cues, and if the whole sample is neutralized by a nonreinforced trial, then groups with the higher proportions of nonreinforced trials should show higher learning rates. If a part of the sample is neutralized and a part connected to the alternative response, then groups with the higher proportions of nonreinforced trials should show lower learning rates.

METHOD

Apparatus.—Each of the two booths in a darkened room contained a response box with two spring-loaded levers and a reinforcing light above each lever. The buzzer signal and reinforcing lights were controlled by a timer, and responses were recorded automatically.

Design.—There were two main parts in the experiment, three groups in each part, and 48 Ss per group. Both parts were run concurrently. In Part I, groups were given 0, 33, or 67% E_0 trials by random combination of 0, 30, or 120 E_0 trials with 60 reinforced trials having an E_1 to E_2 ratio of 50:50. Following this, Ss were immediately shifted to a reinforcement schedule of 90:10 without E_0 events for 100 trials. In Part II, groups were given 0, 33, or 67% E_0 trials by combination of 0, 50, or 200 E_0 trials with 100 reinforced trials having an E_1 to E_2 ratio of 90:10. Sequences of reinforcing and nonreinforcing events were constructed by obtaining 20-trial blocks of reinforced trials and randomly inserting in these appropriate numbers of E_0 trials. Each group had 12 different noncontingent sequences, 4 Ss to a sequence, and sequences were matched wherever possible across all groups. Side of the more frequent reinforcing light was counterbalanced within each sequence subgroup.

Subjects.—The 288 Ss were volunteers from introductory psychology courses at the University of Minnesota.

Procedure.—The standard prediction instructions (Estes & Straughan, 1954) were given with no interpretation for the nonreinforced trials. Practice trials were presented with outcomes in the order $E_1E_1E_2E_2$-$E_0E_0E_1E_2$. If there were questions concerning the E_0 trials, E replied, "That may happen on some trials. In any case, make a choice on each trial."

The sequence of trials was then given without interruptions. The timing of events within trials was as follows: Duration of the signal, 2 sec.; time between the signal and the reinforcing light, 1 sec.; duration of the reinforcing light, 1 sec.; time between the reinforcing light and the next signal, 2 sec.

RESULTS

Part I.—Shown in Fig. 1 are the curves of learning under a reinforcement ratio of 90:10, following three different conditions of preliminary training. Analysis of learning rates was based on the mean number of A_1 responses over the 100 trials, yielding values of 84.58, 83.25, and 80.97 for the groups having 67%, 33%, and 0% nonreinforcements, respectively, during preliminary trials. The linear trend among these means yielded a significant F ($4.97; df = 1/72; P < .05$) in the analysis of variance for response curves (Ostle, 1954).

In order to investigate the effect of E_0 trials on the rate of learning in a more detailed fashion, a statistic, ϕ, was used, defined by:

$$\phi = \frac{\sum_{i=1}^{n'} (p_{i+1} - p_i)}{\sum_{i=1}^{n'} (1 - p_i)}$$

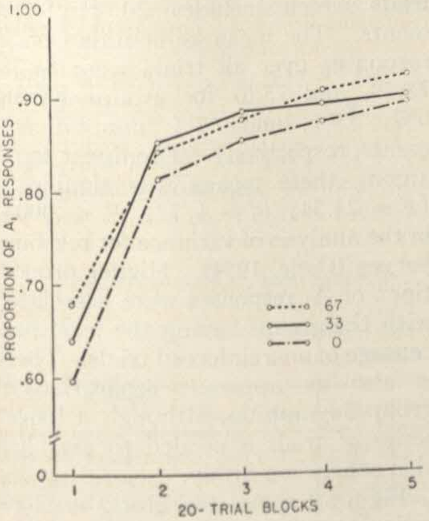


FIG. 1. Learning curves under a 90:10 reinforcement ratio with no E_0 trials, following training under a 50:50 reinforcement ratio with 0%, 33%, or 67% E_0 trials.

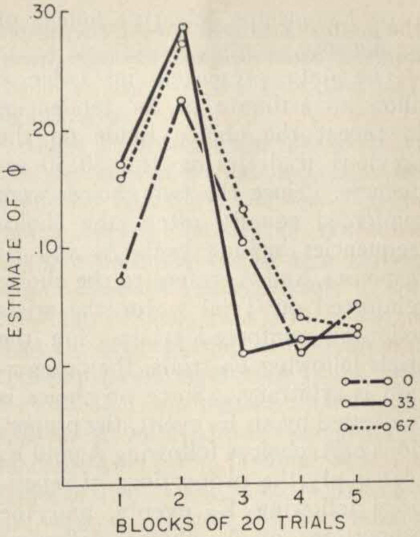


FIG. 2. Block-by-block estimates of ϕ for the 90:10 reinforcement condition of Part I.

counting only the trials on which E_1 was given on Trial i. This statistic estimates θ in the model presented by Estes and Straughan (1954). In the present model, ϕ is greater than θ whenever the proportion of neutral elements is greater than zero. It is therefore predicted that the groups receiving E_0 trials in the pretraining sequence should show higher values of ϕ than the group receiving no E_0 trials. Block-by-block estimates of ϕ for these groups are graphed in Fig. 2. Estimation of ϕ for each S over all trials yielded means of .087, .135, and .163 for the 0%, 33%, and 67% groups, respectively. A variance estimate of .0321 ($df = 72$) was obtained as the residual MS of the factorial analysis of variance, and 95% confidence intervals were estimated as follows:

$$\tfrac{1}{2}(\bar{\phi}_{33} + \bar{\phi}_{67}) - \bar{\phi}_0 = .062 \pm .063;$$
$$\bar{\phi}_{67} - \bar{\phi}_{33} = .028 \pm .073$$

The method of orthogonal contrasts thus revealed a difference of debatable significance between the no-E_0 group and the E_0 groups combined, whereas the difference between the two E_0 groups clearly failed of significance.

The data presented in Table 1 allow an estimate of *S*s' tendencies to repeat the choice made on the previous trial during the 50:50 sequences. Since the two choices were reinforced equally often, the choice frequencies include both A_1 and A_2 responses, and A_i refers to the choice reinforced on Trial *n* for the trials following reinforced trials. For the trials following E_0 trials, the designation is arbitrary. Since no choice is reinforced by an E_0 event, the proportions of A_i choices following A_i and E_0 are simply the proportions of repetitions following E_0 events, and the proportions of A_i choices following A_j and E_0 are the proportions of alternations following E_0 events. In this study, as in the study of Anderson and Grant (1957), *S*s showed a higher proportion of A_i choices following E_0 trials on which they chose A_i than they did following E_0 trials on which they chose A_j. As Table 1 also indicates, *S*s tended to repeat their previous choices on trials following reinforced trials as well.

TABLE 1

PROPORTION OF A_i RESPONSES ON TRIAL $n+1$ UNDER THE 50:50 CONDITION OF PART I

% E_o	Response on Trial *n*	Block 1		Block 2		Block 3	
		E_{in}	E_{on}	E_{in}	E_{on}	E_{in}	E_{on}
0	A_{in}	.50	—	.61	—	.60	—
	A_{jn}	.42	—	.46	—	.53	—
33	A_{in}	.50	.50	.60	.53	.61	.54
	A_{jn}	.42	.50	.55	.47	.52	.46
67	A_{in}	.52	.53	.61	.59	.59	.54
	A_{jn}	.48	.47	.50	.41	.55	.46

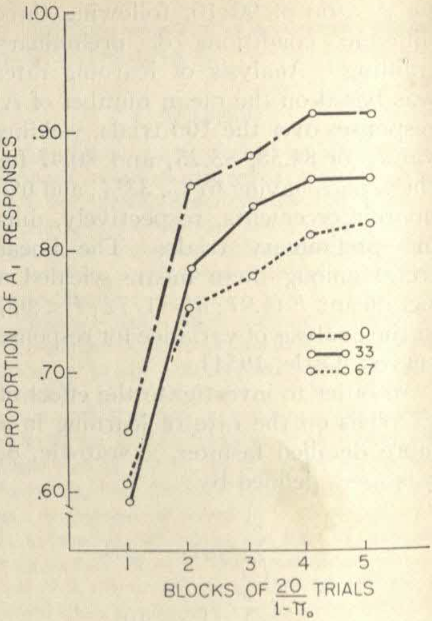


FIG. 3. Learning curves under a 90:10 reinforcement ratio with 0%, 33%, or 67% E_0 trials.

Part II.—The effect of the proportion of nonreinforced trials during learning under a 90:10 reinforcement ratio is shown in Fig. 3. Each point on the curves represents a block of trials which includes 20 reinforcements. The mean percentages of A_1 responses over all trials were 84.38, 78.79, and 75.67 for groups having 0%, 33%, and 67% nonreinforcements, respectively. The linear trend among these means was significant ($F = 24.34$; $df = 1/72$, $P < .0005$) in the analysis of variance for response curves (Ostle, 1954). Higher proportions of A_1 responses were associated with the groups having the least percentage of nonreinforced trials. There is also an apparent separation of group asymptotes, although a longer series of trials is needed to establish this conclusion firmly.

Figure 4 shows the block-by-block estimates of ϕ for these groups. The means for groups receiving 0%, 33%, and 67% E_0 trials were .108, .324, and .247, respectively. The residual *MS*

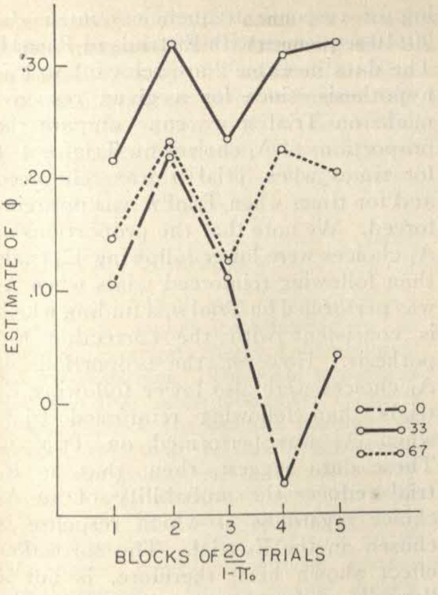


FIG. 4. Block-by-block estimates of ϕ for the 90:10 reinforcement condition of Part II. (The groups were given 0%, 33%, or 67% E_0 trials.)

estimate of σ^2 from the factorial analysis of variance was .0546 ($df=72$) and 95% confidence intervals were estimated as follows:

$$\tfrac{1}{2}(\bar{\phi}_{33} + \bar{\phi}_{67}) - \phi_0 = .177 \pm .082;$$
$$\bar{\phi}_{67} - \bar{\phi}_{33} = -.077 \pm .096$$

Thus, orthogonal contrasts revealed a significant difference between the no-E_0 group and the E_0 groups combined, whereas the difference between the two E_0 groups again failed of significance.

Table 2 contains the results of an analysis similar to that presented in Table 1. However, the lack of symmetry under the 90:10 reinforcement schedules requires a somewhat different presentation. The columns labeled "E_{in}" in Table 2 contain proportions of choices on trials following reinforced trials, with E_1 and E_2 trials pooled. The partition involving A_{1n} and A_{2n} does not, of course, involve pooling over the responses on Trial *n* as is the case in Table 1. Since the proportions of A_1 choices are consistently higher following A_{1n} than following A_{2n}, it appears that *S*s tended to repeat their previous choices following both E_0 and E_i trials, as they did under 50:50 training of Part I.

DISCUSSION

The main hypothesis under test was that nonreinforcement connects cues to responses other than the choice alternatives in a two-choice learning situation. This hypothesis was supported in the first part of the experiment where groups having a greater proportion of E_0 trials showed higher rates of learning on a

TABLE 2

PROPORTION OF A_1 RESPONSES ON TRIAL $n+1$ UNDER THE 90:10 CONDITION OF PART II

% E_0	Response on Trial *n*	Block 1		Block 2		Block 3		Block 4		Block 5	
		E_{in}	E_{on}	E_{in}	E_{on}	E_{in}	E_{on}	E_{in}	E_{on}	E_{in}	E_{on}
0	A_{1a}	.79	—	.89	—	.91	—	.94	—	.92	—
	A_{2n}	.42	—	.64	—	.70	—	.69	—	.79	—
33	A_{1n}	.78	.51	.88	.75	.92	.80	.92	.81	.91	.83
	A_{2n}	.53	.41	.66	.55	.63	.56	.78	.58	.80	.74
67	A_{1n}	.72	.68	.84	.81	.86	.85	.90	.87	.87	.87
	A_{2n}	.55	.46	.64	.52	.64	.56	.65	.62	.71	.56

subsequent 90:10 reinforcement condition. This result and the theoretical background are strengthened considerably by comparing the present group orderings with those in a related study (LaBerge, 1959b), in which neutral elements were depleted by 50:50 training with no E_0 events. In the present study, groups having longer preliminary training showed faster learning, whereas in the previous study groups having longer preliminary training showed slower learning. Both results are consistent with the neutral elements assumption.

Quantitative analysis in terms of the neutral elements model yields some additional support for the neutralizing hypothesis. If the probability of an E_0 event during the 50:50 trials is .67, and if it is assumed that all the sample cues are neutralized on an E_0 trial, then it is expected by probability matching considerations that the population of cues at the beginning of the 90:10 condition will contain 67% neutral elements. When we compare the 67% E_0 group of Part I with the 0% group of Part II, we note that their curves are virtually identical. An independent estimate of the proportion of neutral elements at the beginning of training for the 0% group of Part II constitutes a check on the neutral elements proportion of .67. The relevant parameters estimated for a similar study (LaBerge, 1959a) were .755 for the initial proportion of neutral elements, and .032 for θ. With these parameter values, and allowing for the practice trials in the present experiment, we obtain .66 for the estimated proportion of neutral elements at the beginning of the 90:10 condition for the 0% group of Part II. Thus it would seem that a schedule of 67% E_0 events maintained the proportion of neutral elements at its original level during the 50:50 schedule.

However, it would seem theoretically reckless to let the close quantitative agreement between these values rule out possible additional processes during E_0 trials. A likely candidate would be a correction factor, namely the strengthening of the response opposite to the one made. This could account for the ordering of response frequencies under the 90:10 sequences with E_0 trials in Part II. The data in Table 2 are relevant to this hypothesis, since for a given response made on Trial n we can compare the proportions of A_1 choices on Trial $n + 1$ for times when Trial n was reinforced and for times when Trial n was not reinforced. We note that the proportions of A_1 choices were lower following E_0 trials than following reinforced trials when A_1 was performed on Trial n, a finding which is consistent with the correction hypothesis. However, the proportions of A_1 choices were also lower following E_0 trials than following reinforced trials when A_2 was performed on Trial n. These data suggest, then, that an E_0 trial reduces the probability of an A_1 choice regardless of which response is chosen on the E_0 trial. The correction effect shown here, therefore, is not a "pure" one, for it apparently occurs only when the more frequent response is made on an E_0 trial.

Another way of evaluating the evidence for a general correction effect consists of examining the differences among the groups shown in Fig. 3. If E_0 did not affect response probabilities, as expected from an identity operation, then the three group curves should be identical, since the groups differed only with respect to the presence of E_0 trials which were added to equivalent series of reinforced trials. The fact that instead we obtained a lowering of A_1 proportions with increasing proportions of E_0 trials argues for some sort of average correction effect, while seriously questioning the validity of the identity hypothesis.

Since slopes of response probability curves are rather insensitive measures of the proportion of neutral elements in the population, it is expected that even a small correction factor in the E_0 process would easily mask any effect on learning rate due to neutralizing. Thus, it should be no surprise to find that increased proportions of E_0 trials during the 90:10 condition lowered the level of A_1 responding as shown in Fig. 3. If a neutralization effect did occur over these trials, it should be revealed by the ϕ curves of Fig. 4, which measure rate of learning due to E_1 reinforcements. The slopes of these curves are apparently less steep for groups with E_0 trials, which suggests that rate of learning for these groups is higher for a longer time, compared with the group having no E_0 trials.

SUMMARY

Six groups of 48 *S*s each were trained in a noncontingent two-choice situation in an attempt to investigate effects of nonreinforcement upon learning. The main hypothesis tested was that nonreinforcement connects at least part of the sampled elements to responses other than the choice alternatives. Three of the groups were given preliminary trials under a 50:50 reinforcement ratio with 0%, 33%, or 67% of the trials nonreinforced. These groups were then shifted to a 90:10 reinforcement condition without further nonreinforcement. Groups having greater amounts of nonreinforcement during preliminary training showed significantly higher rates of learning under the 90:10 condition, supporting the neutralization hypothesis.

For further analysis of the neutralization effect and in order to test the identity hypothesis, 0%, 33%, or 67% of nonreinforced trials were inserted in a 90:10 reinforcement schedule. Groups having more nonreinforced trials showed lower proportions of A_1 responses, indicating that nonreinforcement is not in general a neutral event, as required by the identity hypothesis. Estimates of learning rates by a conditional probability method supported the neutralization hypothesis for these groups.

REFERENCES

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.

ATKINSON, R. C. An analysis of the effect of nonreinforced trials in terms of statistical learning theory. *J. exp. Psychol.*, 1956, **52**, 28–32.

ESTES, W. K. The statistical approach to learning theory. In S. Koch, (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

LABERGE, D. Effect of preliminary trials on rate of conditioning in a simple prediction situation. *J. exp. Psychol.*, 1959, **57**, 20–24. (a)

LABERGE, D. A model with neutral elements. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. (b)

NEIMARK, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 249–252.

OSTLE, B. *Statistics in research.* Ames, Iowa: Iowa State Coll., Press, 1954.

(Received December 6, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 214–215

SUPPLEMENTARY REPORT: PARTIAL REINFORCEMENT AND AMOUNT OF REINFORCEMENT AS DETERMINANTS OF INSTRUMENTAL LICKING RATES[1]

STEWART H. HULSE AND W. EDWARD BACON
Johns Hopkins University

Hulse (1958) used large and small wet-mash rewards and examined resistance to extinction of a running response after these amounts had been given on 50% and 100% of the training trials. After partial reinforcement, the large reward produced much faster running than the small reward, but after continuous reinforcement, the large reward produced slower running than the small reward. The present experiment generalizes this observation to instrumentally-conditioned licking, following a suggestion (Hulse, Snyder, & Bacon, 1960) that the phenomenon might hold for licking as well as for running.

Method.—The *S*s were 50 male albino rats. None had prior experience in licking from tubes. The apparatus and the general technique for instrumentally conditioning licking have been described elsewhere (Hulse, 1960; Hulse et al., 1960). In the present experiment, two schedules of reinforcement were used: 8 to 1 fixed ratio (FR8) and continuous reinforcement (CRF). These schedules were combined in a factorial design with five concentrations of a saccharine reinforcer: .25, .50, 1.80, 4.50, and 10.00 gm. of saccharine added to 1 liter of tap water. Each reinforcement consisted of a .005-cc drop of the appropriate concentration of saccharine. The *S*s were tested both hungry and thirsty. Food (7.5 gm.) and water (20 cc minus the amount obtained in the experimental apparatus) were given in the form of wet mash .5 hr. to 1.0 hr. after testing on each day.

On Day 1 of training, all *S*s were permitted 100 licks of the appropriate concentration of saccharine on a CRF schedule. Licking was then shaped such that by Day 7, all *S*s were producing 800 licks on the appropriate schedule of reinforcement. This regime was continued through Day 20. On Day 21, 4 days of extinction were begun. Each day, *S*s were permitted to lick on the dry tube for 5 min.

Results.—During training, daily licking rates were computed from the times *S*s required to emit the first 400 and the last 400 licks. These data, averaged across the last 5 days of training, are shown in Fig. 1. Three general effects (Ps $< .001$ in an analysis of variance) are shown in the figure. First, licking rate increases monotonically as a function of the log of reinforcement concentration. Second, with CRF, rates decrease from the first to the last half of a session, while with partial reinforcement, rates increase across the two blocks of licks. Finally, partial reinforcement yields slower average rates of licking than CRF. This last observation bears further investigation since, with responses other than licking, ratio reinforce-

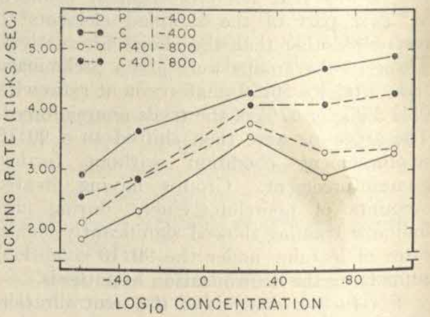


FIG. 1. Acquisition licking rates for FR8 and CRF (P and C, respectively) and five concentrations of saccharine. (Means are shown for the first and last 400 licks.)

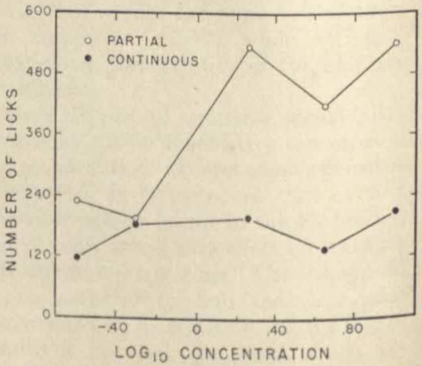


FIG. 2. Licks during extinction following FR8 and CRF with five concentrations of saccharine.

---

[1] This research was supported by National Science Foundation Research Grant G8712 to the senior author.

ment in a free operant situation yields a higher rate of responding than CRF.

During extinction, concentration, schedule, and their interaction are all significant ($Ps < .01$). Figure 2 shows that partial reinforcement produces greater resistance to extinction than CRF. Also, the number of licks emitted during extinction increases as a function of the log of reinforcement concentration. This latter relation is not simple, however, since by *t* test, none of the CRF groups differ significantly from one another ($Ps > .05$). With partial reinforcement, on the other hand, resistance to extinction increases markedly as a function of reinforcement concentration. By *t* test, the means for the three highest concentrations all differ significantly from those for the two lowest concentrations ($Ps < .01$).

REFERENCES

HULSE, S. H., JR. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, 56, 48-57.

HULSE, S. H., JR. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, 3, 1-3.

HULSE, S. H., SNYDER, H. L., & BACON, W. E. Instrumental licking behavior as a function of schedule, volume, and concentration of a saccharine reinforcer. *J. exp. Psychol.*, 1960, 60, 359-364.

(Received February 4, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 2, 215-216

SUPPLEMENTARY REPORT: SHOCK INTENSITY AND UNCONDITIONED RESPONDING IN A SHUTTLE BOX

THOMAS R. TRABASSO[1] AND RICHARD W. THOMPSON

Michigan State University

Kimble (1955) studied UCRs of rats to different intensities of shock. He found that shock elicited two kinds of responses, "flinching" at lower intensities (0.1–0.6 ma.) and "jumping" which increased in frequency as shock intensity increased (0.1–0.9 ma.). However, he used an ascending-descending order of shock presentation and the dimensions of his shockbox (11 in. long, 9 in. high, and 4½ in. wide) differed from those of typical shuttleboxes (24–36 in. long and 5–14 in. high). Since the order of shock presentation may influence the frequency of responding and the dimensions of the shuttlebox have been found to limit the response repertoire of *S* (Denny & Thomas, 1960) the present study was designed to replicate Kimble's with modification of these variables.

Method.—The *S*s were 18 male and 18 female albino rats (Kimble's *S*s were 10 males), ranging in age from 94 to 169 days (Kimble's *S*s were aged 200 days); groups matched for age and sex were randomly assigned to the two experimental conditions (ceiling heights).

The apparatus was a shuttlebox, 36 in. long and 4 in. wide with either a 5 in. or a 14 in. ceiling. The stimulator, timer, adaptation period, shock duration (1 sec.), intershock interval (30 sec.), and intensity range (0.0–0.9 ma. in 0.1 ma. steps) were identical to those used by Kimble. Two sessions of 80 shocks with a 5-min. rest interval between sessions were given. Each of the 10 levels of shock was given eight times in a different random order for each session and each *S*.

The responses were operationally defined: (*a*) No Response; (*b*) Flinch—*S* made a startle response to shock, no more than one paw leaving the grid; (*c*) Shuttle-Compatible—this response class is analogous to Kimble's "Jump" category and was divided, in the present study, into three subclasses Run, Prance—*S* advanced or remained within a 9-in. section with both hind feet leaving the grid nearly simultaneously and alternately raised one of the forefeet, and Jump—all four of *S*'s feet left the grid. The main response classes (*a*, *b*, and *c*) are mutually exclusive but the subclasses of Class *c* are not.

Results and discussion.—For the Flinch and Shuttle-Compatible response classes a separate analysis of variance for each class, using shock levels as repeated measures was performed. For both classes, the only significant effect was attributed to shock intensity (Flinch, $F = 53.53$; Shuttle-Compatible, $F = 293.29$; 8/256 *df*, $P < .01$). Neither ceiling height, sex, nor any interactions were significant in either analysis. The data for the 36 *S*s were summarized for

---

[1] National Science Foundation Cooperative Fellow. The authors wish to thank M. Ray Denny for his helpful advice and criticism throughout the course of this study.

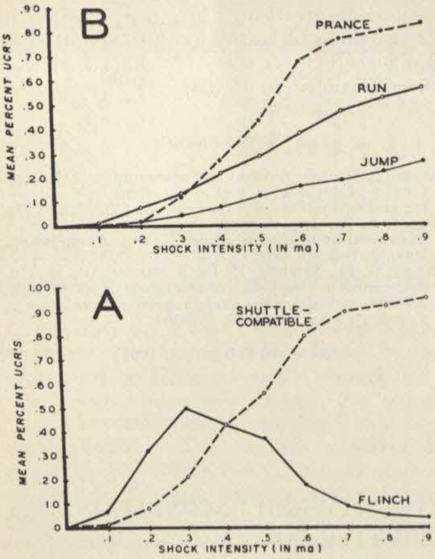


FIG. 1. Mean percentage of *S*s showing (A) Flinch and Shuttle-Compatible, and (B) Prance, Run, and Jump responses as a function of shock intensity.

the Flinch and Shuttle-Compatible classes and are presented as mean percentages of responding as a function of shock intensity in Fig. 1A.

In general, our results conform very well to those reported by Kimble. The *S*s in the present study flinched more frequently at all levels than Kimble's (e.g., Kimble's *S*s flinched 36% at 0.3 ma. vs. 50% in Fig. 1A, at the same point). With respect to the Shuttle-Compatible responses, there were no apparent differences between Kimble's "Jump" data and ours.

Separate analyses of the Shuttle-Compatible subclasses (Run, Prance, and Jump) using the two sessions as repeated measures yielded no significant interactions or differences between sexes or ceiling heights. These data are summarized in Fig. 1B. The percentages reported do not add to 100% since it was possible for an individual *S* to give all three responses to one shock presentation.

Kimble reported a significant difference in median threshold for jumping between first and second halves of his experiment (.36 ma. vs. .51 ma.). There were no significant differences between our medians (.439 ma. vs. .468 ma.; $t = 1.18$, $df = 35$, $P > .05$). Neither did the frequency of the Flinch responses differ between the two sessions ($t = 0.61$). Hence, no evidence of adaptation to shock was observed.

The findings of the present study support the generality of Kimble's results. Presumably, identification of the UCRs to shock is necessary for a thorough understanding of escape and avoidance learning.

REFERENCES

DENNY, M. R., & THOMAS, J. O. Avoidance learning and relearning as a function of shuttlebox dimensions. *Science*, 1960, **132**, 620–621.

KIMBLE, G. A. Shock intensity and avoidance learning. *J. comp. physiol. Psychol.*, 1955, **48**, 281–284.

(Received March 18, 1961)

Journal of
Experimental Psychology

Vol. 63, No. 3 MARCH 1962

EFFECTS OF PATTERN OF REINFORCEMENT AND VERBAL INFORMATION ON ACQUISITION, EXTINCTION, AND SPONTANEOUS RECOVERY OF THE EYELID CR[1]

THOMAS F. HARTMAN AND DAVID A. GRANT

University of Wisconsin

This study was an investigation of the effects of certain patterns of reinforcement on the acquisition, extinction, and spontaneous recovery of the eyelid CR. Following the different patterns of acquisition, the effects of informing *S* that the UCS would not be delivered during extinction and spontaneous recovery were determined. One general purpose of the study was to obtain information as to the degree that verbal information (and presumably self-instruction) may affect extinction and spontaneous recovery in the eyelid CR. Four patterns of reinforcement were investigated: 100% reinforcement, 50% random reinforcement, double-alternation reinforcement, and double-alternation reinforcement with a trial counter and complete knowledge of the sequence of reinforced trials.

Grant, Riopelle, and Hake (1950) found marked inhibition of responding to double alternation of reinforcement when compared with 50% random reinforcement. Since instructions regarding the nature of the task have been found to affect the rate of conditioning with the conditioned eyelid response (e.g., McAllister & McAllister, 1958; Miller, 1940), the conditioned salivary response (Razran, 1935), and the conditioned GSR (Cook & Harris, 1937), it seems logical to assume that information concerning the patterning of reinforcement would be of importance. To test this hypothesis a patterned group that would have complete information about the reinforcement sequence (double alternation given reinforcement "rule" and trial counter) was compared with a patterned reinforcement group, double alternation, that received no such aid in recognizing the pattern of reinforcement.

Norris and Grant (1948) found that instructing *S*s to try to inhibit their eyelid CRs produced slower acquisition and more rapid extinction. Instructions have also been found to

---

[1] This research was supported in part by a grant from the National Science Foundation and in part by the Research Committee of the Graduate School of the University of Wisconsin with funds provided by the Wisconsin Alumni Research Foundation.

affect the extinction of other CRs such as the conditioned GSR (Cook & Harris, 1937) and the conditioned finger withdrawal (Lindley & Moyer, 1961), but the extinction of some responses do not show this effect (Joyce, 1956). As the methodology of the eyelid CR has been more thoroughly explored than those of the GSR, hand-withdrawal response and verbal CRs, study of verbal parameters of this response would be desirable in any event. One methodological complication of the eyelid CR offers an unusual opportunity for investigation of the effects of instructions. Because *S*s in eyelid conditioning experiments can be separated into voluntary and nonvoluntary responders solely on the basis of response form (Hartman & Ross, 1961; Spence & Ross, 1959), possible differential effects of patterns of reinforcement and of verbal information on these two classes of *S*s should prove most interesting.

METHOD

Apparatus.—The apparatus was essentially that described by Hartman and Grant (1960). The CS consisted of an 0.8-mL. increase in brightness of a 10-cm. circular mild glass disk. Ambient room illumination was approximately 1.0 mL. CS duration was 700 msec.; the UCS, a corneal air puff, occurred 540 msec. following the onset of the CS and was of 200 msec. duration. The intensity of the air puff was regulated by a 150-mm. column of mercury.

Procedure.—A 4 × 2 × 2 factorial design was used. The four acquisition conditions were 100% reinforcement (100%), 50% random reinforcement (R), double-alternation reinforcement (DA), and double alternation with a counter under the CS window to count the trials (DAC). Before extinction and before spontaneous recovery half of the *S*s were told that they would not receive the air puff for the remainder of that session. Acquisition, extinction, and spontaneous recovery conditions were orthogonal to each other.

All *S*s received 60 acquisition trials, 30 extinction trials, and a 30-trial spontaneous recovery test approximately 24 hr. later. The intertrial interval varied between 20 and 40 sec. with a mean of 30 sec. The trials were programed by means of a Western Union tape transmitter. The "random" condition was completely random except the last acquisition trial was always reinforced and the first and last halves of the acquisition series contained an equal number of reinforcements. Five random sequences were used.

Before beginning the initial experimental session *S*s were given the "neutral" instructions used at the University of Wisconsin laboratory. Those *S*s in Group DAC were also told what the reinforcement pattern in the session would be, and were quizzed until they could tell whether any given trial would include an air puff. The counter was not operative for the other groups but was in the field of vision. In addition all *S*s were told that they would be interrupted three times during the experimental session; those groups given no information about extinction were interrupted twice "to see if they were comfortable" and once to tell them when the session was half over. The groups given information about extinction were told that the third time they were interrupted, it would be to inform them that for the remainder of that session, they would get no more air puffs. If questioned about this statement *E* said that the meaning would become evident as the session progressed. Interruptions following Trials 5, 45, and 60 were made to minimize the probability of distraction due to the final interruption. On the day *S*s returned for the spontaneous recovery test, appropriate groups were told either that everything would be the same as on the previous day, or that during that session there would be no air puffs.

Subjects.—The *S*s were 112 women and 48 men from the elementary psychology courses. Assignment of *S*s to conditions was random within a replication of the 16 experimental conditions until it became necessary to modify assignment to equalize the number of each sex per group. Five *S*s were discarded: 1 failed to appear for the spontaneous recovery test, 3 adapted to the UCS, and 1 gave a record impossible to score due to an extremely high random blink rate.

RESULTS

All eyelid closures with latencies from 200 to 550 msec. were considered anticipatory responses. In some analyses, specified below, *S*s giving voluntary form responses were removed from the data, using the criterion suggested by Hartman and Ross (1961); i.e., *S*s with time derivatives of eyelid responses to light greater than 35% of the mean derivative of the reflex to the UCS for more than half of their responses were classified as voluntary *S*s.

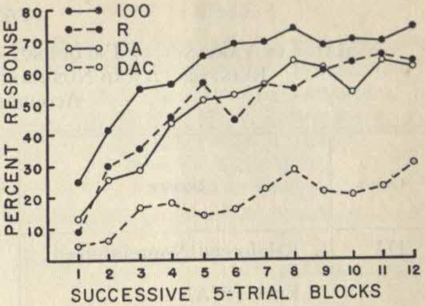


FIG. 1. Acquisition curves under four patterns of reinforcement. (Percentages of CRs are plotted against successive five-trial blocks for 100% reinforcement—100%, 50% random reinforcement—R, double alternation of reinforcement—DA, and double alternation with information and trial counter—DAC.)

Acquisition—Figure 1 shows the mean percentage of anticipatory responses plotted as a function of blocks of five trials for acquisition. In the very first block of trials, the groups no longer appear the same; Group DAC responds at a lower level than the other groups. Group 100% approaches its asymptote faster than the other groups, and the level of response in Group DAC remains very low. The curves of Group R and Group DA virtually overlap during all of acquisition.

The derivative criterion classified 51 *S*s as voluntary: 16 in Group 100%, 14 in Group R, 13 in Group DA, and 8 in Group DAC. A chi square test failed to disclose a significant effect of acquisition series on number of voluntary responders. When these voluntary *S*s were eliminated the same acquisition trends were evident, the only difference being that the overall level of responding was less in all groups. An analysis of variance of the total frequency of responses during the last half of the acquisition session gave highly significant *F*s among reinforcement groups of 19.18 for the data of all *S*s and 13.52 when voluntary *S*s were eliminated, with error estimates based on 156 and 105 *df*, respectively.

Another group of analyses were performed on the terminal acquisition trials. Over the last 28 trials the reinforced and nonreinforced trials in Groups DA and DAC were analyzed. The percentage frequency of responses on reinforced and nonreinforced trials is presented separately for voluntary and nonvoluntary re-

TABLE 1

PERCENTAGE FREQUENCY OF RESPONSES GIVEN BY VOLUNTARY AND NONVOLUNTARY *S*S ON REINFORCED AND ON NONREINFORCED TRIALS DURING THE LAST 28 ACQUISITION TRIALS

Response Type	Group DA			Group DAC		
	N	Reinforced Trials	Nonreinforced Trials	*N*	Reinforced Trials	Nonreinforced Trials
Voluntary	13	73%	67%	8	65%	32%
Nonvoluntary	27	51%	54%	32	22%	15%

TABLE 2

ANALYSES OF VARIANCE OF FREQUENCY OF RESPONSES IN 200–550 MSEC. RANGE ON REINFORCED AND NONREINFORCED TRIALS DURING LAST 28 ACQUISITION TRIALS

Group	Source	Voluntary *S*s		Nonvoluntary *S*s	
		df	*F*	*df*	*F*
DA	Reinforced-Nonreinforced	1	2.25	1	0.80
	*S*s	12	19.36**	26	12.97**
	Error (*MS*)	12	(2.07)	26	(3.32)
DAC	Reinforced-Nonreinforced	1	11.32*	1	2.30
	*S*s	7	1.02	31	4.10**
	Error (*MS*)	7	(7.56)	31	(3.59)

* $P = .05$.
** $P = .01$.

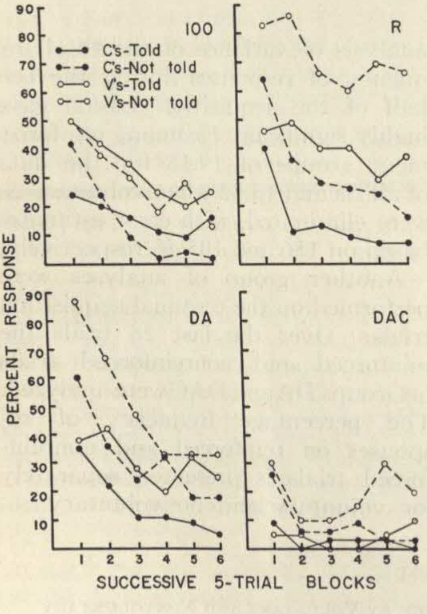


FIG. 2. Extinction of eyelid responses following four patterns of reinforcement during acquisition. (Percentages of CRs are plotted against successive five-trial extinction blocks. Each reinforcement group is subdivided into *S*s giving CRs—Cs—and *S*s giving voluntary responses—Vs—on the basis of the form of the response. These are further subdivided on the basis of whether they were informed—Told—when extinction procedures started or whether they were not so informed—Not Told.)

sponders in Groups DA and DAC in Table 1. There appear to be no systematic differences between responses to reinforced and nonreinforced trials in Group DA. In Group DAC, however, there is less responding on nonreinforced trials, and this is especially evident among the voluntary responders. The summaries of the analyses of variance are given in Table 2 where it is seen that the only significant difference between reinforced and nonreinforced trials was found among the voluntary responders of Group DAC. A notable absence of individual differences also occurred in this subgroup; every one of the 8 voluntary responders gave an equal or greater frequency of responses on reinforced trials than they did on nonreinforced trials. For these *S*s the information about the reinforcement sequence and the aid of the counter definitely affected responsiveness.

Extinction.—Figure 2 presents the extinction curves for the nonvoluntary and voluntary *S*s separately. The ordinate is percentage of responses in the 200- to 550-msec. interval; the abcissa is successive five-trial blocks. Two major effects are evident in these curves: *S*s told when extinction begins give fewer responses at the beginning of extinction than those *S*s not given this information, and those *S*s classified as voluntary responders on the basis of form give more responses in extinction than the nonvoluntary *S*s regardless of the information given them.

Six separate analyses were made of the extinction data. Two analyses utilized the data of all *S*s. The other four analyses considered either nonvoluntary *S*s or voluntary *S*s separately. With these last four analyses those *S*s that did not make five anticipatory responses during the acquisition series were not included in the analysis. It was hoped that this procedure would reduce the error variance and partially compensate for the loss of power due to reduced numbers and unequal numbers per group in these analyses. The procedure used to adjust the bias caused by the unequal numbers of *S*s per group was that outlined by Snedecor (1946, pp. 284–293).

Analyses of variance were performed on the last 30 trials within each acquisition reinforcement group to determine if those groups of *S*s which subsequently received differential extinction and spontaneous recovery treatment were significantly different in acquisition. The largest F found was 2.09 ($df = 3/36$, $P > .05$). Separate analyses were performed for the data of all *S*s, nonvoluntary and voluntary *S*s.

TABLE 3

ANALYSES OF VARIANCE OF FREQUENCY OF RESPONSES IN 200–550 MSEC. RANGE FOR ALL *S*s DURING EXTINCTION

Source	*df*	All Trials *F*	First Five Trials Only *F*
Acquisition (A)	3	8.52**	10.87**
Extinction information (E)	1	3.28	8.31**
A × E	3	1.32	0.88
Error (*MS*)	152	(51.91)	(2.53)

** $P = .01$.

TABLE 4

ANALYSES OF VARIANCE OF THE EFFECT OF EXTINCTION PROCEDURE INFORMATION ON THE FREQUENCY OF RESPONSES IN 200–550 MSEC. RANGE FOR ALL *S*s DURING EXTINCTION IN EACH REINFORCEMENT GROUP

Group	Source	*df*	All Trials *F*	First Five Trials Only *F*
100%	Information	1	0.03	0.35
	Error (*MS*)	38	(63.21)	(2.87)
R	Information	1	4.66*	6.78*
	Error (*MS*)	38	(78.63)	(2.95)
DA	Information	1	0.18	1.13
	Error (*MS*)	38	(55.66)	(3.53)
DAC	Information	1	2.66	5.43*
	Error (*MS*)	38	(10.13)	(0.74)

* $P = .05$.

The analyses of variance of the extinction data of all *S*s are presented in Table 3. When all trials are considered, the extinction information has no significant effect, but when the first five trials only are considered this information produces an effect significant at the .01 level.

Expectancy theory would predict that verbal information about onset of extinction would be most effective with Group R; therefore, it was decided to evaluate the effects of information separately in each acquisition group. These analyses are presented in Table 4. When the data for all trials are considered the effects of the extinction information are evident only with Group R, but if only the first five trials are considered, both Group R and Group DAC show significant effects. In each group the effect of extinction information was to reduce responsiveness, especially on early extinction trials.

To determine if the differences between nonvoluntary *S*s and voluntary *S*s during extinction apparent in Fig. 2 were significant, separate analyses were performed for each response type. These analyses are presented in Table 5. If all trials were considered, extinction information had no significant effect for either group; when the first five trials only were considered, significantly fewer responses were found in the group told when extinction began for both the nonvoluntary and voluntary *S*s. Examination of each acquisition group separately produced only two significant *F*s, both with the voluntary *S*s in Group R. With the data on all trials, the *F* was 5.86, and with only the first five trials, the *F* was 5.98, both significant at the .05 level (error estimate based on 12 *df*). No significant *F* due to extinction information was found with Group DAC, but the test here with the voluntary *S*s was very insensitive as the error estimate was based on only 6 *df*.

TABLE 5

ANALYSES OF VARIANCE OF FREQUENCY OF RESPONSES IN 200–550 MSEC. RANGE DURING EXTINCTION FOR *S*s MEETING ACQUISITION CRITERION

Type of *S*	Source	*df*	All Trials *F*	First Five Trials Only *F*
Non-voluntary	Acquisition (A)	3	2.15	4.33*
	Information on extinction (E)	1	2.43	8.71**
	A × E	3	0.81	0.08
	Error (*MS*)	66	(44.14)	(2.39)
Voluntary	Acquisition (A)	3	5.23**	3.66*
	Information on extinction (E)	1	3.67	6.70*
	A × E	3	1.18	1.40
	Error (*MS*)	41	(61.01)	(2.46)

* $P = .05$.
** $P = .01$.

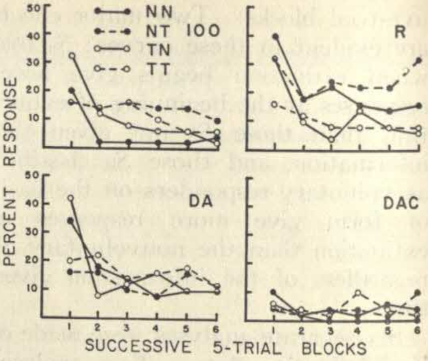


FIG. 3. Percentage frequency of eyelid responses plotted for successive 5-trial blocks during the spontaneous recovery test consisting of 30 extinction trials, following four patterns of reinforcement and a 30-trial extinction 24 hr. earlier. (After each acquisition procedure, *S*s are further subdivided into four groups according to whether they were informed—T—or not informed—N—of the absence of the UCS on initial extinction and on the spontaneous recovery test.)

Spontaneous recovery.—The spontaneous recovery data are plotted as Fig. 3; the axes are as in the previous figures, but the parameter separates groups according to whether they received extinction information (T) or not (N) on initial extinction and on the spontaneous recovery test. Comparing Fig. 3 with Fig. 2 there is evidently considerable spontaneous recovery, and again the effect of information at the beginning of the test session is evident. Those *S*s who were told there will be no air puffs at the beginning of the spontaneous recovery test session always give fewer responses than the corresponding group which receives no information. This is true regardless of whether *S*s had or had not been given extinction information on the occasion of first extinction.

Only two analyses were performed on the data gathered during spontaneous recovery; analyses of variance of all trials and the first five trials only for all *S*s. Considering nonvoluntary and voluntary *S*s separately was not feasible due to the small numbers which would be in some of the groups if the data were further divided. Table 6 contains summaries of the analyses of the spontaneous recovery data for all *S*s, first for the data for all trials then for the first five trials only. As with the extinction data information has an effect only if the data for the first five trials are considered, not for the entire test session.

Discussion

Acquisition.—The outstanding result on the over-all acquisition curves is the lowered responding in Group DAC. In this respect, Group DAC resembles Group DA in an earlier study (Grant et al., 1950). Group DA of the present study gave far more responses than the DA group of the earlier one, and the only possible clue to the discrepancy lies in unsystematically recorded verbal reports of *S*s in the two experiments. Whereas Grant, Riopelle, and Hake (1950) reported a large share of the *S*s being able to report the double-alternation pattern, very few of the *S*s in Group DA of this study could verbalize the double alternation in the reinforcement pattern. As pointed out below, perception of the pattern of reinforcement seems to be an important factor in reducing responsiveness.

TABLE 6

Analyses of Variance of Frequency of Responses in 200–550 Msec. Range for All *S*s during Spontaneous Recovery

Source	*df*	All Trials	First Five Trials Only
		F	*F*
Acquisition (A)	3	4.20**	6.68**
Information on extinction (E)	1	0.53	0.02
Information on spontaneous recovery (SR)	1	0.20	7.47**
A × E	3	1.07	1.03
A × SR	3	0.17	0.62
E × SR	1	0.03	0.35
A × E × SR	3	0.23	0.12
Error (*MS*)	144	(29.96)	(1.77)

** *P* = .01.

In Table 1 were reported the percentages of responses in the CR latency range of *S*s in Groups DA and DAC on reinforced and on nonreinforced trials. Each group was separated into the voluntary responders and the nonvoluntary responders on the basis of the objective criterion reported by Hartman and Ross (1961). Table 1 shows that neither the voluntary nor the nonvoluntary *S*s responded differentially to the pattern reinforcement in Group DA. Furthermore, the general response rate was as high as in Group R. In Group DAC, however, there was differential responding to the reinforced and nonreinforced trials, and it followed a very interesting pattern. Only in the case of the voluntary *S*s on the reinforced trials was the response level as high as that in Group DA. The nonvoluntary *S*s, although they showed a slight but insignificant tendency to follow the reinforced-nonreinforced trial pattern, showed a marked general reduction in responsiveness to both the reinforced and nonreinforced trials. In fact, they were responding at less than 50% of the level attained by their fellow nonvoluntary *S*s in Group DA. The voluntary *S*s responded less than 50% as often on the nonreinforced trials as they did on the reinforced trials, and this result was statistically significant as shown in Table 2. Evidently, only the voluntary *S*s were able to make effective use of the information and cues regarding the double alternation reinforcement patterns. But the tremendous reduction in CRs among the nonvoluntary *S*s in

Group DAC was itself a very important finding. It appears to be a phenomenon over and above that of the general reduction in responsiveness obtained when nonreinforced trials are given as in Groups R or DA. Information as to whether a trial will or will not be reinforced, instead of assisting *S* in responding appropriately, seems to result in a generalized inhibition. That this is more than general distraction of *S* by the trial counter, etc., seems indicated by the fact that the voluntary *S*s are able to respond at a high rate on the reinforced trials in Group DAC. The experiment provides no evidence as to whether the generalized inhibition characterizes the double alternation with a reinforcement cue, or whether it is due to the reinforcement cue alone or to some other factors in the experiment.

Extinction.—The general extinction curves shown in Fig. 2 are similar to those usually obtained when it is borne in mind that an acquisition criterion of five anticipatory responses during the training trials was required for *S* to be included in the analysis. The relatively high response level of Group DA during extinction as compared with Group 100% and Group R indicates that randomness of reinforcement alone was not a factor in the resistance to extinction, although only in the voluntary *S*s of Group R was resistance to extinction found to persist through the entire extinction series. Presumably, Group DA, receiving a patterned reinforcement sequence, should be able to discriminate the onset of extinction more readily than Group R. This evidently was not the case, indicating that the DA *S*s could not use the regularity of the reinforcement pattern to detect the onset of extinction just as they could not use the pattern in responding differentially to reinforced and nonreinforced trials during acquisition.

The effect of extinction information was most pronounced in Group R, and least pronounced in Group 100%. The statistical tests of Table 4 indicate that the extinction information was most effective on the first five trials of extinction except in Group R where a persisting effect of the information could be demonstrated through the entire extinction process. Extinction information seemed to be no more effective for the voluntary *S*s than it was for the nonvoluntary *S*s. Here is an indication that the voluntary *S*s were not especially able to utilize information given as to extinction and to adopt "correct strategy" in their responses during extinction. Instead, they tend to continue responding at a fairly high level throughout the extinction series. They made as much use of the extinction information but no more use of it than did the nonvoluntary *S*s.

Two somewhat contradictory aspects of the responding of the voluntary *S*s seem to be demonstrated. First, during acquisition in Group DAC, the voluntary *S*s were able to respond differentially to the reinforced and nonreinforced trials when given appropriate information. Secondly, during extinction, the appropriate information as to the extinction procedure produced no greater effect on the voluntary *S*s than it did on the nonvoluntary *S*s. Instead, the voluntary *S*s continued to respond at a high level throughout extinction. In this latter aspect, the responsiveness of voluntary *S*s seemed to be characterized by high *D* (Spence, 1958). Unfortunately, because of the extremely rapid extinction in Group DAC, it is not possible to tell whether the same *S*s who responded differentially during reinforcement were able to utilize the extinction information. But it seems clear that under some circumstances, voluntary *S*s can utilize information to alter their responses according to a "correct strategy;" that in other instances, they are no more able to do so than other *S*s. Alternatively, voluntary responders may include some *S*s who tend to "respond appropriately" to stimulus patterns and some *S*s who merely respond at a high level.

Spontaneous recovery.—As usual (Grant, Hunter, & Patel, 1958; Hartman & Grant, 1960; Howat & Grant, 1958;

Porter, 1938, 1939; Prokasy, 1958), considerable spontaneous recovery was obtained on the spontaneous recovery test trials. These responses occurred primarily on the first extinction block and were followed by rapid extinction. The information as to whether reinforcement would be given during the spontaneous test trials was very important on the first five trials of the spontaneous recovery test regardless of whether *S*s had been told or not been told about the extinction process in the first extinction procedure the day before. Evidently, the recovered CRs can still be affected by extinction information after an actual extinction process has been carried out. Furthermore, whether previous information on extinction had been given or not, no persisting effect of the information on first extinction was found during spontaneous recovery, as was shown by the curves in Fig. 3 and by the absence of significant *F*s for extinction information and extinction by spontaneous recovery information interaction in Table 4. It appears that even after a first actual extinction followed by a 24-hr. recovery period, *S*s are still able to respond differentially to the information that no reinforced trials will be given during the spontaneous recovery test. Something survives the first extinction that can be affected by verbal information about the extinction procedure.

SUMMARY

The effects of 100% reinforcement, (Group 100%) 50% random reinforcement, (Group R) double-alternation reinforcement (Group DA), and double-alternation reinforcement with a trial counter (Group DAC) upon the acquisition, extinction, and spontaneous recovery of the conditioned eyelid response were studied. In addition, the effects of telling *S* at the beginning of extinction or spontaneous recovery that the air puff would no longer be presented were observed. All *S*s received 60 acquisition trials, 30 extinction trials, and a 30-trial spontaneous recovery test 24 hr. later. The CS was a light and the UCS was a corneal air puff. The principal findings were as follows:

1. Groups DA and R produce the same degree of conditioning in acquisition, but Group DAC showed a considerably lower level of response than the other two groups reinforced 50% of the time.
2. In Group DA, neither the nonvoluntary nor voluntary *S*s responded significantly more to the reinforced than to the nonreinforced trials; however, in Group DAC, the voluntary *S*s responded significantly more to the reinforced trials.
3. Knowledge of cessation of the UCS produced a significant decrement of response during the first five extinction trials with both nonvoluntary and voluntary *S*s. The voluntary *S*s did not show immediate extinction when told the air puff would be omitted. Voluntary *S*s, therefore, showed two contradictory aspects in their responding. They were able to respond significantly more to the reinforced trials in Group DAC, but they were not able to utilize knowledge of the cessation of the UCS during extinction to respond appropriately. Instead, they responded at a generally higher rate than the nonvoluntary *S*s during extinction.
4. Considerable spontaneous recovery was found on the first five-trial block followed by rapid extinction. Knowledge that the UCS would not be given on the spontaneous recovery test reduced the number of responses on the test regardless of the occurrence of an earlier actual extinction series with or without verbal instructions as to the cessation of the UCS on the earlier extinction trials.

REFERENCES

COOK, S. W., & HARRIS, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21**, 202–210.

GRANT, D. A., HUNTER, H. G., & PATEL, A. S. Spontaneous recovery of the conditioned eyelid response. *J. gen. Psychol.*, 1958, **59**, 135–141.

GRANT, D. A., RIOPELLE, A. J., & HAKE, H. W. Resistance to extinction and the pattern of reinforcement. I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40**, 53–60.

HARTMAN, T. F., & GRANT, D. A. Effect of intermittent reinforcement on acquisition, extinction, and spontaneous recovery of the conditioned eyelid response. *J. exp. Psychol.*, 1960, **60**, 89–96.

HARTMAN, T. F., & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 334–338.

Howat, M. G., & Grant, D. A. Influence of intertrial interval during extinction on spontaneous recovery of conditioned eyelid responses. *J. exp. Psychol.*, 1958, **56**, 11–15.

Joyce, J. N. The role of language as a cue in learning and extinction. *Dissertation Abstr.*, 1956.

Lindley, R. H., & Moyer, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *J. exp. Psychol.*, 1961, **61**, 82–88.

McAllister, W. R., & McAllister, D. E. Effect of knowledge of conditioning upon eyelid conditioning. *J. exp. Psychol.*, 1958, **55**, 579–583.

Miller, J. The effect of facilitatory and inhibitory attitudes on eyelid conditioning In E. R. Hilgard & D. G. Marquis (Eds.), *Conditioning and learning.* New York: Appleton-Century, 1940.

Norris, E. B., & Grant, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *Amer. J. Psychol.*, 1948, **61**, 37–49.

Porter, J. M., Jr. The modification of conditioned eyelid responses by successive series of nonreinforced elicitations. *J. gen. Psychol.*, 1938, **19**, 307–323.

Porter, J. M., Jr. Experimental extinction as a function of the interval between successive nonreinforced elicitations. *J. gen. Psychol.*, 1939, **20**, 109–134.

Prokasy, W. F. Extinction and spontaneous recovery of conditioned eyelid responses as a function of amount of acquisition and extinction training. *J. exp. Psychol.*, 1958, **56**, 319–324.

Razran, G. H. S. Conditioned responses: An experimental study and a theoretical analysis. *Arch. Psychol., N. Y.*, 1935, **28**, No. 191.

Snedecor, G. W. *Statistical method.* (4th ed.) Ames, Iowa: Iowa State Coll. Press, 1946.

Spence, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

Spence, K. W., & Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

(Received December 27, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 227–234

PATTERN RECOGNITION OVER DISTORTIONS, BY HUMAN SUBJECTS AND BY A COMPUTER SIMULATION OF A MODEL FOR HUMAN FORM PERCEPTION

LEONARD UHR
University of Michigan
CHARLES VOSSLER
System Development Corporation
AND JAMES ULEMAN
Harvard University

The recognition of single randomly generated, "meaningless" patterns has been studied during the past few years by a number of investigators (Attneave, 1957; Crook, Gray, Hanson, & Weisz, 1959; Vanderplas & Garvin, 1959b). Several experiments have further examined the abilities of human *S*s in recognizing such patterns over certain simple variations, such as random noise (Crook et al., 1959; Hillix, 1960), contour noise (Fitts & Leonard, 1957), and systematically continued transformations (LaBerge & Lawrence, 1957). But no experiments have been reported in which *S*s were asked to learn sets of variants of a pattern through experience with individual examples of these sets. Yet this is the typical procedure for computer "pattern recognition" simulation programs (Selfridge & Neisser, 1960; Uhr, 1962). It also presents a good experimental paradigm for the learning of patterns by human beings. Given *N* patterns and *n* variants of each pattern, *S* is presented with each of the *nN* particular pattern instances, asked to give it the correct name of the *N* names, and then told the correct name. This, in more systematic form, is what the child experiences during the natural experiments of his day-by-day learning. That is, the child is given particular instances of such things as "chair," "dog," the letter "A," or his mother's face; and he learns to build up general concepts, or pattern classes.

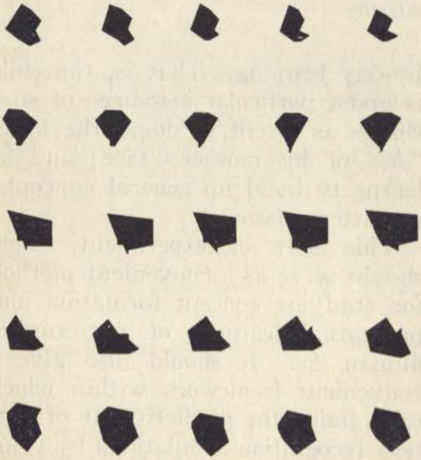
This sort of experiment, then, should serve as a convenient method for studying concept formation and perceptual learning of patterns in human *S*s. It should also give a convenient framework within which to examine the predictive fit of pattern recognition simulations by computer programs, when they purport to be embodiments of theoretical models of human form perception. The use of meaningless patterns for stimulus sets reduces the wealth of information that the human can bring to bear. It also allows for control and systematic variation of the parameters of the stimulus sets.

The main experiment examined the learning of pattern sets as a function of (*a*) interrelations between patterns, (*b*) practice, and (*c*) individual differences between *S*s (including the simulated model as an *S*). Subsidiary experiments examined the effects of different (*a*) durations of stimulus presentation, (*b*) size of pattern set, and (*c*) complexity of pattern.

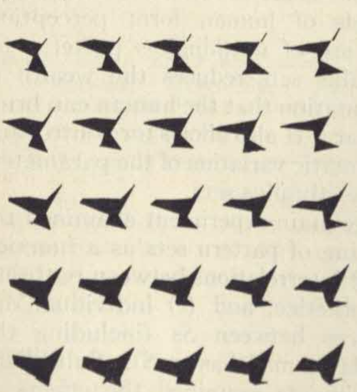
In addition, three form perception

experiments were replicated using the computer as *S*. These were those of Attneave (1957) and Vanderplas and Garvin (1959b), who examined *S*s' abilities to make fine discriminations between individual patterns, and Fitts and Leonard (1957), who examined *S*s' abilities to generalize from individual patterns to variants.

The computer program is the first version of a model for form perception and concept formation. The program does not appear to take advantage of such trivially superior abilities of a digital computer as perfect rote memory or extreme arithmetic precision. Nor does it have built-in imprecision to simulate so-called imprecisions in human *S*s, such as memory decay. Finally, it should be noted that this is not a trivial test for computer programs of this sort, at the present stage of their development. In fact, relatively few programs are capable of even attempting to handle arbitrary new pattern arrays, and it is not known whether any would do appreciably better than the present program (Highleyman, 1961; Uhr, 1962).



1a. The "easy" set (taken from Crook et al., 1959).



1b. The "hard" set (taken from LaBerge & Lawrence, 1957).

FIG. 1. Stimuli used to make up the two experimental sets of five variants of each of five different patterns.

METHOD

Main Experiment

Subjects and design.—Human *S*s were tested in groups of 6 to 10, approximately half men and half women. For the analyses of variance, 3 men and 3 women were chosen randomly from each group. The *S*s were all paid volunteer college students. Each *S* was tested on two stimulus sets; half were given the "unrelated" stimulus set first, the other half were given the "interrelated" stimulus set first. A test on a stimulus set consisted of five learning trials through the entire set of 25 stimuli, i.e., five variants of each of the five patterns. The order of presentation was determined by using a random number table, with ordering different for each of the five trials.

Apparatus and stimuli.—Two sets of patterns were chosen that would be sufficiently difficult so that *S*s could not learn them immediately, but sufficiently learnable so that reasonably interesting learning curves would be generated. These stimulus sets are reproduced in Fig. 1.[1] Each set consisted of five pattern types, each pattern type with five variants. One set was chosen from Crook's eight-sided patterns (Fig. 1a). This was the "easy," "unrelated" set. The second set was chosen from LaBerge and Lawrence's

---

[1] We would like to thank M. N. Crook, of Tufts University, and D. H. Lawrence, of Stanford University, for their kindness in supplying us with usable copies of their original stimuli.

(1957) systematically varied patterns, in which the five different pattern types are actually related to one another (Fig. 1b). This was the interrelated or hard set. The "distance" as measured in an objective manner among members of the *same* pattern type (and more or less confirmed by the subjective judgments collected by LaBerge and Lawrence) is in certain cases *greater than* the distance *between* members of *different* pattern types.

Stimuli were projected, one at a time, by means of a reflector projector, onto a screen approximately 5 to 8 ft. in front of *S*s. Stimulus size was approximately 3 in. square, which subtended roughly a .5° to 1° arc at the retina. These distances were chosen to to give optimal viewing conditions and still be comparable to the computer's resolving power. Pretesting was used to determine the most favorable lighting and viewing conditions, ones that would result in maximal learning for human *S*s. Each stimulus was presented for 5 sec. Ten to 15 sec. after its removal, *S*s were told what the correct name or type was. The following stimulus appeared 3 to 7 sec. after that.

Several subsidiary experiments were run as checks on some of these control conditions. (*a*) One subsidiary control group was given the same stimulus set under the same conditions, except for a 2-sec. presentation period. (*b*) An additional group of *S*s was given a set of seven patterns, including the five that had been used for the "easy" set, in order to see to what extent short-term memory for too large a number of different pattern sets might be serving as a limiting factor for learning. (*c*) The effects of different complexities of stimuli were examined with a third group of *S*s, who were given four-sided patterns. (Complexity was defined by the number of sides to the randomly generated stimulus.) In each of these subsidiary experiments, all other conditions were held constant. Results are presented as comparisons between the two pertinent groups involved.

Finally, a computer program designed to recognize patterns across their variants in a manner suggested by psychological and physiological data (Uhr & Vossler, 1961, 1962) was also tested on these same input patterns. Inputs were presented to the computer in the same order that they had been presented to *S*s. The resolution of the inputs was also the same. Although there was no possibility of making a meaningful equation between computer program and *S*s for such matters as duration of presentation of stimuli, illumination, and other viewing conditions, an effort was made to give *S*s the best possible viewing conditions.

Briefly, the pattern recognition program works as follows: Unknown patterns are presented to the computer in discrete form, as a 20 × 20 matrix of zeros and ones. The program randomly generates measuring operators that simulate local neuron nets, by one of several random methods, and uses this set of operators to transform the unknown input matrix into a list of four 3-bit characteristics. These characteristics are then compared with lists of characteristics in memory, one for each type of pattern previously processed. As a result of similarity tests, the name of the list most similar to the list of characteristics just computed is chosen as the name of the input pattern. The characteristics are then examined by the program and depending on whether they individually contributed to success or failure in identifying the input, amplifiers for each of these characteristics are then turned up or down. This adjustment of amplifiers leads eventually to discarding operators which produce poor characteristics, as indicated by low amplifier settings, and to their replacement by newly generated operators.

Replications

All replication experiments used only the computer as *S*. Here again an attempt was made to equate computer and human conditions wherever possible. Attneave (1957) asked *S*s to discriminate among members of a set of eight randomly produced variants of a prototype. The computer was tested on one of the several sets which Attneave found essentially equivalent.[2] Vanderplas and Garvin (1959b) asked *S*s to learn a set of eight low-association-value patterns in a paired-associate procedure, in which the other member of the association was a nonsense syllable. Fitts and Leonard (1957) had *S*s learn 12 different prototype patterns during three 6 to 10 min. self-study periods.[3] The 70% who passed the criterion of three out of four correct test trials were then tested

---

[2] The authors are grateful to F. Attneave for supplying us with his original stimuli.

[3] Fitts and Leonard (1957, p. 30) have presented results of their experiment in this mode, and P. Fitts has kindly supplied the authors with the original stimuli, along with details (Leonard & Fitts, unpublished manuscript) of their experiment.

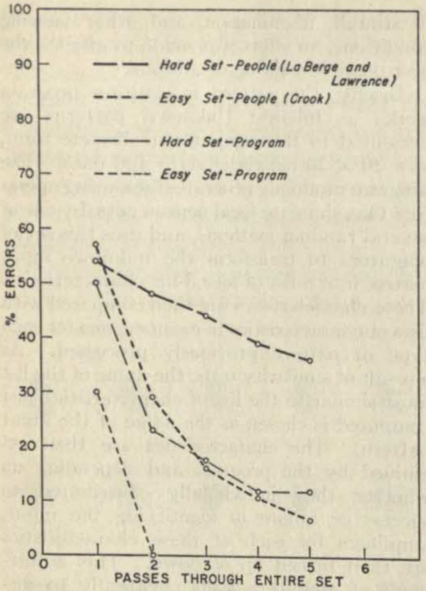


FIG. 2. Mean percentage errors for human *S*s (from 6 to 10 *S*s per point) compared with percentage errors for computer simulation program, on the easy and on the hard sets of stimuli (first presentation).

on generalization to new variants of the prototypes.

RESULTS

Main Experiment

Figure 2 presents mean errors on successive trials ("passes") for *S*s and for the computer simulation program. Table 1 presents results from the analysis of variance on these data. The two pattern sets, and their two presentations, have been analyzed as a single dimension of four "conditions," because of limitations in available computer programs. The program's performance can probably most meaningfully be compared with the human performance in learning the *first* of the two sets learned, since the program does not accumulate broad general abilities and facility for new problems through experience in any way comparable to the human being. Therefore, it did not appear worthwhile to run a "second set" experiment using the program as *S*. Instead, Set 1 results were used as best estimates of what performance would have been on Set 2. (This would appear to be a conservative estimate, since it is saying in effect that no learning will occur, although, in fact, the program *is* capable of self-modification, and actually "learns" to recognize patterns originally through a process of this sort.) Differences between "conditions," between "passes," and between *S*s were all significant beyond the .001 level.

TABLE 1

ANALYSIS OF VARIANCE OF ERRORS AS A FUNCTION OF PATTERN SETS, TRIALS, AND *S*S

Source	Simulation Program as Contrasted with Mean Human Performance			Simulation Program as One of 7 *S*s, along with Individual Human *S*s		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *S*s	1	1265.63	16.40**	6	91.75	7.34***
Trials (T)	4	1163.21	15.08***	4	207.62	34.85***
Conditions (C)	3	1562.63	20.26***	3	539.05	43.16***
*S*s × T	4	36.31	.47	24	5.69	.96
*S*s × C	3	207.42	2.69	18	12.49	2.10*
T × C	12	36.81	.48	12	17.54	2.95**
*S*s × T × C	12	77.11		72	5.94	
Total	39			139		

Note.—The computer simulation program is included as an *S*.
* $P = .05$.
** $P = .005$.
*** $P = .001$.

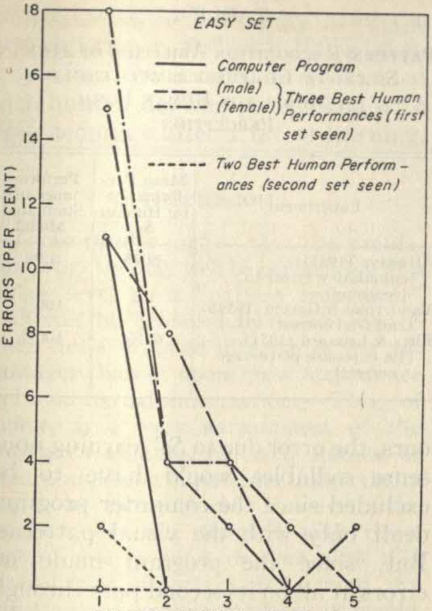


FIG. 3. Computer simulation program performance compared with best individual human performances on the easy set of stimuli.

In particular, the difference between the simulation program and the mean human performance was significant beyond the .005 level, the simulation's performance being consistently better.

Figure 3 presents results, for the unrelated pattern set, for the best of the individual *S*s, and for the computer program. On this set of patterns, some of the human *S*s when they had already had the experience provided by the other set of patterns, outperformed the simulation program. Figure 4 presents results, for the interrelated pattern set, for the best of the individual *S*s, and for the program. On this more difficult set of patterns, the best human performances do not exceed the performance of the program. Nor, in fact, does the best human performance appear to improve very markedly as a result of the additional experience afforded *S*s doing this task as their second task. That is, most of the improvement in the overall group performance appears to result from the improvement of the poorer members of the group.

Figures 5, 6, and 7 present results, for human *S*s alone, for (*a*) different durations of presentation of stimuli—for 2-sec. and 5-sec. durations; (*b*) different size of pattern sets—for five and for seven patterns; and (*c*) different pattern complexities—four-sided and eight-sided patterns. There were no significant effects attributable to either number of pattern sets or duration of stimulus presentation (as varied in this experiment). Complexity of patterns did, however, tend to affect recognition, as indicated by the interaction effect (significant at the .05 level) between complexity and *S*s.

Table 2 presents results of the

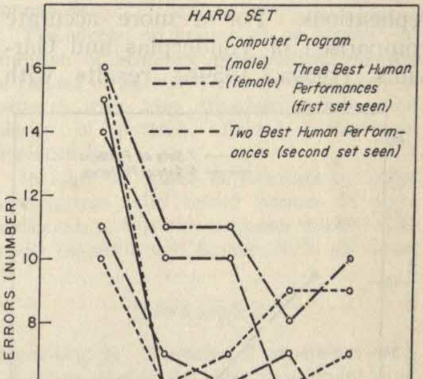


FIG. 4. Computer simulation program performance compared with best individual human performances on the hard set of stimuli.

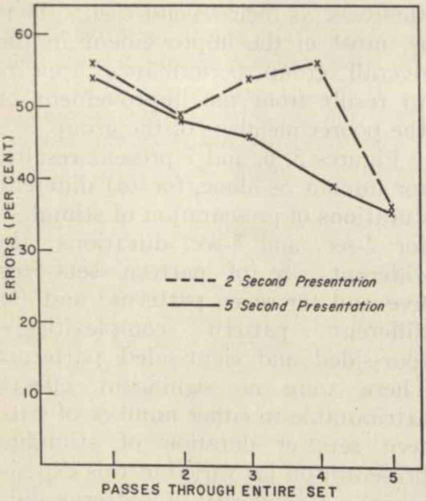


FIG. 5. Effects of duration of stimulus presentation. (Dotted line presents results—10 *S*s—in learning when patterns were presented for a 2-sec. period. Solid line presents results—9 *S*s—when patterns were presented for a 5-sec. period, all other conditions comparable.)

replications. For a more accurate comparison of Vanderplas and Garvin's (1959a, 1959b) results with ours, the error due to *S*s' learning nonsense syllables would have to be excluded since the computer program dealt only with the visual patterns. But, since the program made no errors at all on its second pass through the set, we have not bothered to make such an adjustment. On the Fitts and Leonard (1957) set, the program learned the prototypes to a 75% level after its first experience, and to a 100% level after its second. It gen-

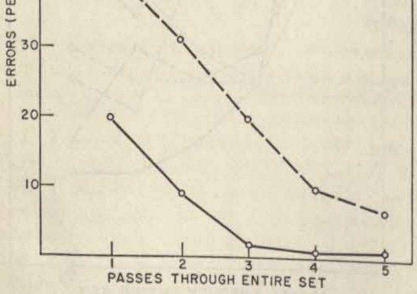


FIG. 6. Effects of size of pattern set. (Dotted line presents results—mean scores for 10 *S*s—in learning seven sets of patterns. Solid line presents results—9 *S*s—for five sets of patterns, all other conditions comparable.)

TABLE 2

PATTERN RECOGNITION ABILITIES OF HUMAN *S*S AND OF COMPUTER SIMULATION OF A MODEL FOR HUMAN FORM PERCEPTION

Experiment	Mean Performance for Human *S*s	Performance of Simulated Model
Attneave (1957) (cumulative errors to learning)	29.75	3.00
Vanderplas & Garvin (1959b) (2nd trial correct)	3%	100%
Fitts & Leonard (1957) (1st exposure percentage correct)	62%	100%

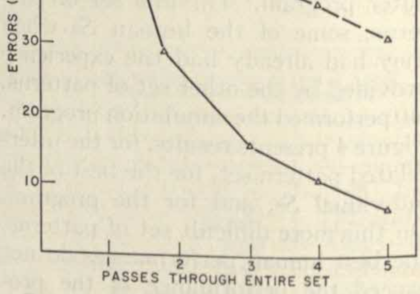


FIG. 7. Effects of number of sides to a pattern. (Dotted line presents results—8 *S*s—in learning four-sided patterns. Solid line presents results—6 *S*s—for eight-sided patterns, all other conditions comparable.)

eralized with 76% accuracy after the first experience, and with 100% accuracy after the second, as contrasted with human *S*s, who generalized with 62% accuracy after 1 hr. of learning.

DISCUSSION

These results suggest that the simulation program may well be performing at a higher level as a "pattern recognizer" than can be achieved by the human *S* when he is divested of some of his advantages based upon past experience and contextual information. This, of course, is a very narrow test of the program's abilities, nor is it known to what population of patterns these results might be generalized. However, the program was not designed with these pattern sets in mind, or even with this general type of pattern set in mind. At this stage of experimentation, it would seem reasonable to simply continue to test such programs out on as wide a range of stimuli as possible.

The fact that the program performs better than human *S*s in this experiment should not be taken very seriously. The simulation is capable of coping with only a part of the perceptual problem. It would not be at all unlikely, or uninteresting, to find that, as the program's generality increases, its abilities at any particular task decreases as a direct and intrinsic function of its greater powers in other directions.

Further experiments of this sort should serve at least two purposes. First, by varying such things as the number of patterns, the number of variants of each pattern, the distance between these variants, the way the individual stimuli sample the pattern's total population, and the distance between patterns, a rather rich domain of behavior closely related to what we have traditionally named "perceptual learning" and "concept formation" should be thrown open to assessment. By testing performance on new variants, different from those originally learned, generalization of classes and concepts can be studied. Second, by comparing the simulation program's performance with human performance as a function of such variations, a beginning can be made in attempting to fit some of the details of the curves generated by the simulated model to the curves generated by human *S*s.

SUMMARY

Human *S*s, and a computer simulation program of a model for form perception, were examined for their behavior in learning to respond with the proper name for a pattern type over variant examples of the pattern. In the main experiment, five pattern types of five variants each were presented to all *S*s (including the simulation program). Four "conditions" were examined: (*a*) unrelated patterns, (*b*) interrelated patterns; each presented as the first, and as the second, set. Differences between *S*s, and between "passes" through the set (learning trials) were also examined. All of the above variables showed differences significant beyond the .001 level. In particular, the simulation of the model performed at a significantly lower error rate (beyond the .005 level) than did the group of human *S*s.

The effects, on human *S*s, or varying (*a*) duration of stimulus presentation, (*b*) size of the set of patterns, and (*c*) complexity of patterns were also examined. Only complexity of pattern affected performance significantly.

In addition, three experiments by other investigators who tested human *S*s were replicated, to test the computer model. The model outperformed human *S*s in all cases.

REFERENCES

ATTNEAVE, F. Transfer of experience with a class-schema to identification-learning of patterns and shapes. *J. exp. Psychol.*, 1957, **54**, 81–88.

CROOK, M. N., GRAY, F. E., HANSON, J. A., & WEISZ, A. The effect of noise on the perception of forms in electro-visual display systems: A set of irregular forms. Tufts University, Inst. Appl. Exper. Psychol., 1959.

FITTS, P. M., & LEONARD, A. J. Stimulus correlates of visual pattern recognition: A probability approach. Laboratory of Aviation Psychology, Ohio State University, 1957.

HIGHLEYMAN, W. H. A note on optimum pattern recognition systems. *IRE Trans. Electron. Comp.*, 1961, No. EC-10, 287–288.

HILLIX, W. A. Visual pattern identification as a function of fill and distortion. *J. exp. Psychol.*, 1960, **59**, 192–197.

LABERGE, D. L., & LAWRENCE, D. H. Two methods for generating matrices of forms of graded similarity. *J. Psychol.*, 1957, **43**, 77–100.

SELFRIDGE, O. G., & NEISSER, U. Pattern recognition by machines and men. *Scient. American*, 1960, **203**, 60–68.

UHR, L. "Pattern recognition" computers as models for form perception. *Psychol. Bull.*, 1962, in press.

UHR, L., & VOSSLER, C. A pattern recognition program that generates, evaluates and adjusts its own operators. *Proc. Western Joint Computer Conf.* 1961, **19**, 555–561.

UHR, L., & VOSSLER, C. A computer simulation of pattern perception and concept formation. In, *Proceedings of the bionics symposium: 1961.* Ithaca, N. Y.: 1962, in press.

VANDERPLAS, J. M., & GARVIN, E. A. The association value of random shapes. *J. exp. Psychol.*, 1959, **57**, 147–154. (a)

VANDERPLAS, J. M., & GARVIN, E. A. Complexity, association value, and practice as factors in shape recognition following paired-associate training. *J. exp. Psychol.*, 1959, **57**, 155–163. (b)

(Early publication received June 16, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 235–243

ON THE HETEROGENEITY OF STIMULUS AND RESPONSE ELEMENTS IN THE PROCESSING OF INFORMATION[1]

WILLIAM C. HOWELL
Ohio State University

Although it has been assumed that "... one or both of two simultaneous performances will show some impairment" (Woodworth & Schlosberg, 1955, p. 88), exceptions to this generalization appear to be many and varied. Klemmer (1956, 1958), for example, has noted that such an impairment is contingent upon task complexity or difficulty. Other investigators have pointed out that performance efficiency in multiple activities is a function of factors such as sense channel differences (Mowbray, 1953), task repetitiveness (Bahrick, Noble, & Fitts, 1954), input redundancy (Broadbent, 1958), and combined task training (Bahrick et al., 1954; Briggs & Waters, 1958). Finally, industrial research has indicated that a certain amount of alternation among tasks, particularly widely differing tasks, is actually beneficial to worker output (Ghiselli & Brown, 1948, pp. 420–423).

The number of variables contributing to the divergence of results in multiple-task studies is probably quite large. None is more basic, however, than the failure of investigators to distinguish between two very different multiple-task situations. One is the *time-sharing* task arrangement in which multiple stimuli, each with its associated response, occur at the same point in time (Bahrick et al., 1954; Briggs & Howell, 1959; Mowbray, 1953). The other involves only the alternation, in some fashion, of stimuli or S-R units defining the tasks and hence may be termed *task heterogeneity* (Baker, Wylie, & Gagné, 1951; Klemmer, 1956, 1958). It is not inconceivable that such dissimilar situations would lead to conflicting results. Temporal coincidence (i.e., time sharing) would virtually assure some decrement in one of the tasks, whereas alternation (i.e., task heterogeneity) could occur without impairment to either.

The present research was concerned with the influence of one of these variables, task heterogeneity, upon simple information-processing performance. Heterogeneity was provided by two stimulus and two response classes, the former distinguished on the basis of sense modality (audition vs. vision) and the latter on the basis of muscle group (verbal vs. manual response mode). Combining these elements in various ways, it was possible to obtain the following types of heterogeneity: (*a*) homogeneity (or zero heterogeneity) in which one stimulus class is paired with only one response class; (*b*) stimulus heterogeneity in which the two stimulus classes are paired with a single response class; (*c*) response heterogeneity in which the two response classes are paired with a single stimulus class; and (*d*) task hetero-

---

[1] This research was carried out in the Laboratory of Aviation Psychology and was supported by the United States Air Force under Contract No. AF 33(616)-3612, monitored by the Aerospace Medical Division. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

geneity in which both stimulus classes are used, each paired with its own response class. The plan for evaluating heterogeneity effects involved comparisons of the efficacy of specific S-R pairs to convey information when included in a heterogeneous task with their efficacy in this regard when included in a homogeneous task. Since task heterogeneity may be considered a combination of heterogeneous stimulus and response elements, a further objective was an evaluation of the relative contribution of stimulus and response heterogeneity to the over-all task heterogeneity effect.

A distinguishing feature of the current approach is that the method for producing heterogeneity (of stimuli, of responses, or of the task) created no change in the amount of information presented. In the typical time-sharing or heterogeneity investigation, the amount of stimulus information increases as tasks are added, so that performance with respect to the individual components may decrease while over-all information transmitted remains constant or even increases. In the present design, however, any changes in information transmitted must be attributed to heterogeneity as defined.

EXPERIMENT I

The first experiment was to establish base performance values for each of four tasks: (*a*) visual stimulus with manual response, (*b*) auditory stimulus with manual response, (*c*) visual stimulus with verbal response, and (*d*) auditory stimulus with verbal response. It should be recognized that such base values, obtained under homogeneous task conditions, were necessary both as a means of evaluating S-R compatibility effects within specific subtasks and as a basis of comparison for the heterogeneous tasks undertaken in the subsequent experiments.

Method

Apparatus and stimuli.—The OSU serial discrimeter was employed to present the stimuli and to score responses. A description of this device has been given elsewhere (Alluisi & Muller, 1958). Functionally, the apparatus permitted the preprograming of six stimulus events in any order desired for a sequence of 100 presentations. The *S* responded to each of these stimuli by pressing one of six response keys (manual response condition) or by pronouncing one of six numbers to activate a voice key (verbal response condition). The activation of either a manual or a verbal response key initiated the presentation of the succeeding stimulus in the sequence with the result that *S* maintained complete control of stimulus pacing. In addition, the activation of each manual key was recorded on one of 36 counters arranged in a 6 × 6 stimulus-response matrix. This scoring arrangement enabled *E* to record directly total errors, as well as individual response frequencies for use in computing entropy values. For verbal response trials it was necessary for *E* to record each response manually.

With the presentation of the first stimulus in each trial, a 1/1000-min. timer was activated which continued to run until the last response in the sequence had been completed. Thus, it was possible to record the over-all time required for the completion of all 100 responses in a sequence and from this to compute the average response time per stimulus presentation.

The six visual and six auditory stimuli used in the study are described fully in Table 1. The visual stimuli were selected from a list of colors demonstrated by Halsey and Chapanis (1951) to be easily discriminable. The stimuli were projected from the rear onto the center of a 10-in. diameter opal glass screen by means of a multiple-unit optical projector. The viewing distance for *S* was approximately 28 in., with the diameter of each color target subtending 31′ of visual angle at the eye.

The six tones used as auditory signals were selected on the basis of psychophysical judgments such that (*a*) the middle frequency bisected the distance in pitch between the two apparatus-imposed extremes, (*b*) the high loudness level was well below the discomfort threshold, (*c*) the low loudness level

TABLE 1

VISUAL AND AUDITORY STIMULI AND ACCOMPANYING RESPONSES: EXP. I

Stimuli				Responses	
Visual		Auditory			
Color Name	Wratten Filter	Frequency (cps)	Intensity (db.)	Verbal	Manual[a]
Light Red	24	100	60	1	LIII
Dark Blue	45a	500	53	2	LII
Light Green	52	6000	58	3	LI
Orange	106	100	90	4	RI
Violet	34a	500	85	5	RII
Gray	ND/0.6	6000	93	6	RIII

[a] Designations refer to the digit numbers (e.g., LIII means third digit on the left hand) and their associated keys.

was well above absolute threshold, and (*d*) loudness levels were matched at all frequencies. All tones were produced by individually preset electronic signal generators (Eico Model 377), the outputs of which were fed into a set of Telephonics headphones worn by *S*.

Subjects.—A total of 36 male students served in this experiment, each of whom was assigned at random to one of four groups. Each group, therefore, contained 9 *S*s, all of whom were trained to perform under one of the four homogeneous S-R conditions: visual-verbal (V-V), visual-manual (V-M), auditory-verbal (A-V), and auditory-manual (A-M).

Procedure.—In this, as well as in the succeeding experiments, all *S*s received the benefit of 6 practice trials (600 S-R pairings) prior to the performance on 3 experimental trials (300 S-R pairings). The amount of training provided was sufficient to bring all *S*s to a comparable level of performance. That this level was probably short of asymptotic is suggested by findings which demonstrate that improvement on such tasks continues for at least 50 100-stimulus trials (Alluisi & Martin, 1957). In both the training and experimental trials *S* was instructed to respond as rapidly and accurately as possible. Each training trial was followed by a demonstration of the S-R pairings to prevent *S* from establishing incorrect associations. The training trials were administered in blocks of three with the same random sequence being repeated throughout each block. Experimental trials, on the other hand, were administered singly with each sequence drawn at random from a population of 27 such sequences. Rest periods of 3-min. duration were interposed between experimental trials and between training blocks; periods of 1 min. were inserted between training trials within blocks. All training and experimental trials were completed in a single session, the duration of which was approximately 1 hr.

Results

Performance under each of the four conditions was evaluated using two information measures: information transmitted (H_t) and information transmitted per second (H_t/sec). Since errors were few and well distributed, the former measure consti-

TABLE 2

MEAN INFORMATION TRANSMITTED (H_t) PER RESPONSE AND RATE OF INFORMATION TRANSMITTED (H_t/SEC) SCORES OBTAINED UNDER THE FOUR HOMOGENEOUS CONDITIONS OF EXP. I

S-R Pairing	H_t (Bits)[a]		H_t/Sec	
	Mean	*SD*	Mean	*SD*
A-V	2.50	.05	1.68	.28
V-V	2.49	.04	1.94	.18
A-M	2.44	.11	2.06	.26
V-M	2.47	.01	2.48	.13

[a] Maximum possible H_t was 2.58 bits per response.

tutes a reasonable index of response accuracy, whereas the latter reflects both speed and accuracy of performance.

Considering first the H_t measure, it is apparent from Table 2 that only small differences occurred among S-R pairings with regard to accuracy. In fact, in no case did the differences exceed 2.4% of the absolute amount of information transmitted. An analysis of variance performed on these data confirmed the lack of statistical significance of the differences.

Turning to the H_t/sec measure, the results clearly indicate the superiority of the manual over the verbal response mode, and of the visual over the auditory stimulus mode. The statistical significance of the differences between means shown in Table 2 was demonstrated using a nonparametric technique developed by Alluisi (1956) for the purpose of testing analysis of variance hypotheses. Use of a nonparametric analysis was prompted by a failure of the H_t/sec data to meet the homogeneity assumption of the parametric analysis of variance. This analysis yielded χ^2 of 7.111 ($df = 1, P < .01$) for Stimulus Class, 7.111 ($df = 1, P < .01$) for Response Class, .445 ($df = 1$, *ns*) for S × R, and 14.667 ($df = 3$) for Total.

It is interesting to note that despite the significance of these major effects, the S × R interaction does not approach significance. Thus, it may be assumed that the S-R compatibility of the four tasks is comparable and hence can safely be ignored when these tasks are combined into various heterogeneity conditions as in the succeeding experiments. It must be added that the common level of S-R compatibility is not particularly high, for the transmission rates never exceeded 2.5 bits/sec.

EXPERIMENT II

Experiment II was concerned with the evaluation of task heterogeneity as it affects information-processing performance. Heterogeneity was introduced into the task scheme by the simple means of combining, as subtasks, equal portions of the homogeneous tasks described in Exp. I. Thus, two distinct heterogeneous tasks were available: visual-verbal, auditory-manual (VV-AM); and visual-manual, auditory-verbal (VM-AV).

The purpose of this experiment was to evaluate heterogeneity effects through comparison of performance on the heterogeneous tasks with performance on the component homogeneous tasks. The latter values, of course, were available from the results of Exp. I. Thus, it was intended that the average scores for each pair of subtasks obtained under homogeneous task conditions serve as the basis of comparison for the same subtasks performed under heterogeneous task conditions. Any differences obtained between these scores would then be attributable solely to heterogeneity effects.

Method

The apparatus and procedures used were the same as those described in Exp. I. Only two groups of *S*s were required for this experiment, one for each of the two task heterogeneity conditions (VV-AM and VM-AV). As in Exp. I, 9 male students were assigned at random to each group.

The specific S-R pairings used in the task heterogeneity conditions are summarized in Table 3. It may be noted that for both Conditions VV-AM and VM-AV the pairings were identical to those employed in Exp. I; that is, three S-R pairs were selected from each of the homogeneous tasks and combined to form two six-alternative heterogeneous tasks. In neither of these heterogeneous tasks was there any deviation from the stimulus sequencing arrangement used in Exp. I. Hence, auditory and visual stimuli

TABLE 3

HETEROGENEOUS STIMULUS-RESPONSE PAIRS USED IN EXP. II

Group I (VM-AV)		Group II (VV-AM)	
Stimuli	Responses	Stimuli	Responses
100 cps, 60 db.	1 (verbal)	Light red	1 (verbal)
500 cps, 53 db.	2 (verbal)	Dark blue	2 (verbal)
6000 cps, 58 db.	3 (verbal)	Light green	3 (verbal)
Orange	RI (manual)	100 cps, 90 db.	RI (manual)
Violet	RII (manual)	500 cps, 85 db.	RII (manual)
Gray	RIII (manual)	6000 cps, 93 db.	RIII (manual)

(calling for verbal and manual responses) were randomly interspersed within each trial sequence. In no case were S-R elements permitted to overlap in time.

Results

As may be seen in Table 4, the differences in H_t/sec obtained between the two heterogeneity conditions were small: the average rate of transmission values were 2.38 bits/sec for Cond. VV-AM and 2.26 bits/sec for Cond. VM-AV. The corresponding average values of H_t for the two conditions also were not dissimilar: 2.51 bits for VV-AM and 2.43 bits for VM-AV. Tests of these differences by the t statistic confirmed their failure to achieve significance.

In view of this lack of significance, the data from the two task heterogeneity conditions were averaged for comparison with mean homogeneous performance (obtained by averaging the scores in Table 2). The difference between mean H_t/sec scores for the homogeneous and heterogeneous task conditions was .29 bits/sec, and a Mann-Whitney U test applied to this difference achieved statistical significance ($P < .01$). Inspection of the accuracy data revealed that both heterogeneous and homogeneous conditions yielded identical mean H_t scores of 2.47 bits.

EXPERIMENT III

In view of the significant influence of task heterogeneity on rate of information transmission, it became meaningful to investigate the relative importance of stimulus and response factors in determining the over-all effect. The purpose of Exp. III, then, was to manipulate stimulus and response heterogeneity independently, comparing each to the basic conditions of Exp. I as well as to the task heterogeneity results of Exp. II.

Method

As in the task heterogeneity conditions (Exp. II), there were available two distinct tasks for testing both the stimulus and response heterogeneity effects. For stimulus heterogeneity these were auditory visual-verbal (AV-V) and auditory visual-manual (AV-M) S-R combinations; for the response heterogeneity conditions they were auditory-

TABLE 4

MEAN H_t AND H_t/SEC SCORES OBTAINED UNDER THE TWO HETEROGENEOUS TASK CONDITIONS

Cond.	H_t (Bits)		H_t/Sec	
	Mean	*SD*	Mean	*SD*
VV-AM	2.51	.08	2.38	.17
VM-AV	2.43	.04	2.26	.17
Mean	2.47	.06	2.32	.17

TABLE 5

MEAN H_t AND H_t/SEC SCORES OBTAINED UNDER STIMULUS AND RESPONSE HETEROGENEITY CONDITIONS

Task Components	H_t		H_t/Sec	
	Mean	*SD*	Mean	*SD*
Stimulus Heterogeneity				
AV-V	2.46	.07	1.90	.25
AV-M	2.48	.08	2.54	.15
Mean	2.47	.08	2.22	.20
Response Heterogeneity				
V-MV	2.49	.11	2.21	.31
A-MV	2.45	.10	2.12	.41
Mean	2.47	.10	2.16	.36

manual verbal (A-MV) and visual-manual verbal (V-MV) combinations.

The apparatus and procedures used were identical to those used in Exp. I and II. Again, 9 male students served as *S*s for each condition. In Cond. AV-V and AV-M the stimuli used were 100, 500, and 6000 cps tones and orange, violet, and gray lights. The verbal responses to these stimuli were 1–6, respectively, and the manual responses LIII to RIII, respectively, as listed in Table 1. In Cond. A-MV and V-MV the Responses 1, 2, 3, RI, RII, and RIII were paired with all six auditory or all six visual stimuli as listed in Table 1.

Results

Mean H_t and H_t/sec scores for both stimulus and response heterogeneity conditions are given in Table 5. It is obvious that no reliable H_t differences occurred among task components, and, further, that these scores are no different from those obtained in Exp. I and II (all values approximate 2.47 bits). The difference between mean H_t/sec scores on the two stimulus heterogeneity conditions was large, the values being 1.90 bits/sec and 2.54 bits/sec for AV-V and AV-M, respectively. A *t* test applied to this difference indicated that it was statistically significant ($P < .01$). On the other hand, the difference between the two response heterogeneity means was not sufficiently large to approach statistical significance ($t = .53$). As a result, the two response heterogeneity conditions (V-MV and A-MV) were combined in their comparison with homogeneous conditions, whereas the two stimulus heterogeneity conditions (AV-V and AV-M) were considered independently.

The stimulus and response heterogeneity effects were again evaluated through comparison with homogeneous base conditions: mean scores on the heterogeneous tasks were compared with mean scores for the component subtasks performed under homogeneous conditions (Table 2). Of the two heterogeneous stimulus tasks only AV-M yielded a score appreciably above the homogeneous counterpart: the difference was .28 bits/sec, significant ($P < .01$) in terms of both *t* and *U* tests. Verbal performance, although insignificantly ($P > .10$) superior under the heterogeneous stimulus condition, was poor for both homogeneous and heterogeneous tasks (1.80 and 1.90 bits/sec, respectively). Turning to response heterogeneity, the mean of 2.16 bits/sec was only .13 bits/sec greater than that for the homogeneous base condition. Both *t* and *U* tests confirmed the insignificance of this difference ($P > .10$).

Comparisons were also made between average stimulus or response heterogeneity performance and that obtained for task heterogeneity in Exp. II. Although neither of the component heterogeneity conditions (i.e., stimulus or response) yielded H_t/sec values as high as those obtained under task heterogeneity, the differences were not statistically reliable.

DISCUSSION

The results clearly indicate that task heterogeneity is beneficial to simple information-processing performance under conditions of low S-R compatibility and relatively little training. The findings with regard to the relative contribution of stimulus and response heterogeneity to the over-all task heterogeneity effect may also be stated simply: stimulus heterogeneity appears to increase information transmission significantly over the homogeneous control condition, whereas response heterogeneity leads to no reliable improvement. Whether the stimulus component is sufficient to account for the complete task heterogeneity effect is not clearly shown by the data. Statistically, there is no significant difference between the H_t/sec values for stimulus and task heterogeneity ($P > .05$), although the mean difference is fairly large in favor of task heterogeneity (.10 bits/sec). Further, the coding scheme employed in one of the stimulus heterogeneity conditions (AV-M) was such that it provided a useful cue for left- vs. right-hand responses. The relatively high transmission rates obtained under this condition may thus be somewhat spurious. It is probable, then, that the heterogeneity of stimuli is one factor contributing to the high rate of transmission associated with task heterogeneity. More conclusive, however, is the indication that stimulus heterogeneity is not the *only* contributing factor.

That such findings appear to be in distinct disagreement with many of the other studies encountered in this area may be attributed partially to differences in emphasis and design as suggested above. However, two investigations bear a sufficient degree of similarity to the present work to warrant further comparison. Klemmer (1956) reported that when *S* is forced to alternate regularly between visual and auditory signals more rapidly than once every 2 sec., his performance on a key pressing task declines sharply. In a later study, Klemmer (1958) was able to demonstrate the same effect for conditions of random alternation between visual and auditory channels.

To what, then, can be attributed the marked difference between the present results and those reported by Klemmer? The most obvious explanation appears to lie in the nature of the S-R pairs: whereas in the present research each response was associated with a single stimulus under any given condition, Klemmer used only three response elements for all six stimuli (three visual and three auditory). As a result, each response was associated with one visual and one auditory stimulus. It appears, then, as if a change in signal mode in some way combined briefly the tendencies for error on one channel with the comparable tendencies for error on the other channel without appreciably strengthening the association of either set of signals with the correct responses. In the present situation, of course, error tendencies would have little opportunity to combine since each signal was associated with its own response. Such an explanation, however, does not appear sufficiently strong to account completely for the large differences observed between the two studies.

Another explanation, also limited, is based upon the fact that Klemmer used forced-pacing of stimuli, whereas the present investigation permitted self-pacing by *S*. Of the two modes of presentation, self-pacing has been shown to yield a higher level of performance (Alluisi & Muller, 1958), although a full description of forced- vs. self-pacing differences has yet to be advanced. Considering the difference between the Klemmer results and those of the present investigations, then, it appears that the independence of S-R units and the manner in which stimuli are paced may determine the extent as well as the direction of the heterogeneity effect: for independent, self-paced S-R pairs task heterogeneity leads to improved performance whereas for nonindependent, forced-paced S-R pairs it incurs a decrement.

Assuming that the above explanation is sufficient to account for the differences reported, the problem of interpreting the beneficial heterogeneity effect obtained in the present research still remains. Why, to state the problem somewhat differently, should a simple task be more difficult to perform than a slightly more complex one?

Reasonable answers may be found within the traditional areas of learning and motivation. The learning interpretation relies upon the concept of stimulus generalization: since the introduction of stimulus or task heterogeneity divides the stimulus population into two distinct classes, the extent of stimulus generalization with respect to each signal should be reduced. As a result, it would be expected that the strength of incorrect or generalized associations (which associations control error scores as well as latency of responding) would likewise be reduced. The motivation hypothesis is less precisely put since it makes use of concepts which are, at best, poorly understood. Briefly stated, stimulus change of the sort occasioned by stimulus or task heterogeneity may reduce the monotony of the task. At a physiological level, this may be equivalent to suggesting that stimulus change increases the general activation level of the nervous system. In any event, such a motivational effect would be most apparent in the latency data, which expectation is borne out in both Exp. II and Exp. III findings. A second point in favor of the motivation hypothesis is the fact that stimulus generalization, upon which the alternative hypothesis is based, could hardly be expected to be very strong in the present situation, even under homogeneous conditions. Indeed, stimuli were chosen for ease of discrimination, and were then differentiated further by six training trials. Certainly, whatever generalization gradients remained would be extremely peaked. It would appear, therefore, that the influence of stimulus heterogeneity on performance is predominantly a motivational one.

Two additional results are of interest in that they bear upon the findings of earlier investigations. First, the visual stimulus mode was superior to the auditory mode with respect to the rate of information transmitted. This is in clear agreement with the conclusions reached by Klemmer (1956) and Mowbray (1953). Second, the manual response mode led to significantly higher transmission rates than did the verbal mode, a result in strict disagreement with the findings of several investigators (Alluisi & Martin, 1957; Alluisi & Muller, 1958). One explanation of this discrepancy has been stated earlier: the present coding scheme favored the manual mode, at least in the stimulus heterogeneity condition, through the provision of left- vs. right-hand response cues. However, superiority of the manual mode was not limited to the stimulus heterogeneity condition; it occurred under the homogeneous control tasks of Exp. I as well. Any cuing available under the latter conditions would also characterize the Alluisi research, since both were similar in the assignment of stimuli to the two hands. The discrepancy in results must therefore be attributed to those features upon which the studies differed, namely, amount of stimulus and response information and S-R compatibility. In regard to the latter variable, it is unfortunate that differences were not obtained between the various subtasks with respect to S-R compatibility. Had such differences appeared, it would have been meaningful to evaluate the influence of S-R compatibility upon the obtained heterogeneity effects.

In conclusion, then, it appears that task performance may be enhanced by the presence, in certain amounts, of stimulus or task heterogeneity. It is to be expected that the heterogeneity effect is not unlimited: the most effective level probably depends upon other task variables such as pacing mode, speed stress, stimulus entropy, S-R compatibility, level of practice, and the like. However, the present series of experiments serve to emphasize that in selecting task conditions the major concern should be with finding *optimum* rather than *minimum* heterogeneity levels.

SUMMARY

Three experiments were carried out to determine the influence of stimulus (S), response (R), and S-R or task heterogeneity on simple information-processing performance. Heterogeneity was provided by two dissimilar S classes (auditory vs. visual) and two dissimilar R classes (manual vs. verbal). The conditions were the types of heterogeneity noted above plus a set of homogeneous control conditions. Under all conditions there were 6 alternative Ss and 6 associated Rs; 300 serial S-R occurrences comprised each task. The major performance index was average rate of information transmitted (H_t/sec). Eighteen *S*s served under each of the three heterogeneity conditions and 9 served under each of the four homogeneous control conditions.

The major results were: (*a*) For the particular stimulus and response classes used there appeared no evidence for S-R compatibility differences, and compatibility generally was low. (*b*) In support of earlier findings, the visual code was superior to the auditory code; contrary to earlier findings, the manual response mode was superior to the verbal response mode. (*c*) Task heterogeneity led to significantly higher rates of information transmission than did task homogeneity. (*d*) The task heterogeneity effect may be attributed partially to the S component, since S heterogeneity also enhanced performance significantly over the homogeneous controls, whereas R heterogeneity did not. Neither S nor R heterogeneity scores differed significantly from task heterogeneity scores.

From these findings it is concluded that simple information-processing performance may be enhanced by the presence, in certain amounts, of stimulus or task heterogeneity. The most effective heterogeneity level for any particular task, however, is probably determined by a number of other task variables.

REFERENCES

ALLUISI, E. A. Computational formulae for a distribution-free test of analysis of variance hypotheses. *USAF WADC tech. Rep.*, 1956, No. 56-339.

ALLUISI, E. A., & MARTIN, H. B. Comparative information-handling performance with symbolic and conventional arabic numerals: Verbal and motor responses. *USAF WADC tech. Rep.*, 1957, No. 57-196.

ALLUISI, E. A., & MULLER, P. F., JR. Verbal and motor responses to seven visual codes: A study in S-R compatibility. *J. exp. Psychol.*, 1958, **55**, 247-254.

BAHRICK, H. P., NOBLE, M., & FITTS, P. M. Extra-task performance as a measure of learning a primary task. *J. exp. Psychol.*, 1954, **48**, 298-302.

BAKER, K. W., WYLIE, R. C., & GAGNÉ, R. M. The effects of an interfering task on the learning of a complex motor skill. *J. exp. Psychol.*, 1951, **41**, 1-9.

BRIGGS, G. E., & HOWELL, W. C. On the relative importance of time sharing at central and peripheral levels. *USN Train. Dev. Cent. tech. Rep.*, 1959, No. 508-2.

BRIGGS, G. E., & WATERS, L. K. Training and transfer as a function of component interaction. *J. exp. Psychol.*, 1958, **53**, 379-387.

BROADBENT, D. E. *Perception and communication.* New York: Pergamon, 1958.

GHISELLI, E. E., & BROWN, C. W. *Personnel and industrial psychology.* New York: McGraw-Hill, 1948.

HALSEY, R. M., & CHAPANIS, A. On the number of absolutely identifiable spectral hues. *J. Opt. Soc. Amer.*, 1951, **41**, 1051-1058.

KLEMMER, E. T. Time sharing between auditory and visual channels. In G. Finch and F. Cameron (Eds.), *Air force human engineering, personnel, and training research.* (NAS-NRC Publ. No. 455). Washington: National Academy of Sciences-National Research Council, 1956. Pp. 199-203.

KLEMMER, E. T. Time sharing between frequency-coded auditory and visual channels. *J. exp. Psychol.*, 1958, **55**, 229-235.

MOWBRAY, G. H. Simultaneous vision and audition: The comprehension of prose passages with varying levels of difficulty. *J. exp. Psychol.*, 1953, **46**, 365-372.

WOODWORTH, R. S., & SCHLOSBERG, H. Experimental psychology. (Rev. ed.) New York: Holt, 1955.

(Received November 16, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 244–247

SPACE PERCEPTION AMONG UNILATERALLY PARALYZED CHILDREN AND ADOLESCENTS [1]

HOWARD T. BLANE

Massachusetts General Hospital, Boston

The sensory-tonic field theory of perception (Werner & Wapner, 1949, 1952) assumes that the degree of muscular involvement is an important variable in perceptual experience. Utilizing this assumption, previously reported studies (Carini, 1955; Goldman, 1953; Wapner, Werner, & Chandler, 1951; Wapner, Werner, & Morant, 1951; Werner, Wapner, & Chandler, 1951) demonstrate that experimentally induced changes in the muscular state of *S* affect the visual perception of verticality and the autokinetic phenomenon. Further, in a wedding of sensory-tonic with Werner's developmental theory (Werner, 1948), Wapner and Werner (1957) have shown a developmental ordering of such effects in the perception of verticality from childhood through adulthood. In keeping with the assumption of the importance of muscular involvement as well as with previously reported findings, two specific hypotheses regarding the visual organization of space may be formulated: (*a*) Enduring asymmetric muscular imbalance in the organism will affect perceptual spatial organization, and (*b*) such spatial organization will vary with developmental level.

PROCEDURE

Four groups of *S*s were selected to participate in two experiments chosen to test the hypotheses advanced above.

Subjects.—Forty male and female postpoliomyelitic adolescents and children were divided into four equal groups: (*a*) children, 9–12 yr. of age (mean = 10.3 yr.), with residual paralysis of the right leg; (*b*) children, 9–12 yr. of age (mean = 11.1 yr.), with residual paralysis of the left leg; (*c*) adolescents, 15–19 yr. of age (mean = 16.5 yr.), with residual paralysis of the right leg; and (*d*) adolescents, 15–19 yr. of age (mean = 16.5 yr.), with residual paralysis of the left leg. The *S*s were selected from a group of 400. postpoliomyelitic children and adolescents being evaluated for skeletal growth at Children's Hospital, Boston, and care was taken to include no case with known psychiatric disturbance, visual defect, or social or mental retardation. Each patient was given a muscle test rating by a physiotherapist, on a scale ranging from 0 (complete paralysis of the affected leg) to 120 (no paralysis).[2] Since scores of 100 or more reflect only barely discernible paralysis, no patient with a score over 100 was included in the final sample. Analysis of variance revealed no significant differences among the four groups with regard to degree of paralysis.

The age at which *S* contracted poliomyelitis and the period of time elapsed between onset and participation in this investigation may be assumed to be variables of consequence, especially from a developmental viewpoint, in *S*'s perceptual functioning. However, to control these variables and simultaneously control the extent of residual paralysis, on the assumption that the latter is a more critical variable, was not possible. Consequently, child and adolescent *S*s differ significantly with respect to age at onset ($F = 10.11$, $df = 1$, $P < .01$), the children being younger at onset (mean = 3.2 yr.) than the adolescents (mean = 5.9 yr.). Similarly, elapsed time since onset is significant ($F = 15.35$, $df = 1$, $P < .01$), the children having had paralysis for an average of 7.5 yr., the adolescents 10.6 yr. The groups show no significant difference, however, with respect to number of corrective orthopedic operations or corrective orthopedic appliances (braces, crutches, canes).

---

[1] This investigation was supported by USPHS Grant M-348 from the National Institute of Mental Health, United States Public Health Service, and was carried out when the author was a graduate student at Clark University. Actual testing of *S*s was done at Children's Hospital, Boston.

[2] I am indebted to Margaret Anderson and her staff at the Growth Study for their assistance in the selection and ratings of *S*s.

Experiment I.—Visual perception of verticality under conditions of body tilt and starting position was tested first. The apparatus used, an adjustable luminescent rod and a tilting chair, as well as the physical conditions of testing, have been described by Werner, Wapner, and Chandler (1951). Two variables were systematically studied: body tilt (B)—30° left (L), erect (E), 30° right (R); and starting position of the rod (S)—30° L, 10° L, 10° R, 30° R. All combinations of these variables were employed, making a total of 12 conditions: (*a*) B, 30° R-S, 30° L; (*b*) B, 30° R-S, 10° L; (*c*) B, 30° R-S, 10° R; (*d*) B, 30° R-S, 30° R; (*e*) B, E-S, 30° L; (*f*) B, E-S, 10° L; (*g*) B, E-S, 10° R; (*h*) B, E-S, 30° R; (*i*) B, 30° L-S, 30° L; (*j*) B, 30° L-S, 10° L; (*k*) B, 30° L-S, 10° R; (*l*) B, 30° L-S, 30° R. Each *S* in the four groups was tested under each of the 12 conditions, which were presented in random order. Under all conditions of testing, *S* instructed *E* as to how to move the rod until it appeared vertical to *S*.

The measure employed was the deviation in half degree of tilt from objective, or plumb line vertical. An angular position to the right of objective vertical (as viewed by *S*) was arbitrarily designated as +, and to the left of objective vertical, as −. The data were treated by analysis of variance designed to evaluate independent groups with repeated measurements.

Experiment II.—The autokinetic phenomenon was the second means chosen to test the above hypotheses. The stimulus object was a pinpoint source of light placed 10 ft. directly in front of the seated *S*. Movement of *S* was minimized by seating him in a chair with side supports and an adjustable head rest. The task for *S* was to report when the light began to move, the direction of movement, and any change of direction of movement. The *S* reported for 35 sec. after he first perceived movement. Three electric timers, activated by microswitches operated by *E*, recorded *S*'s reports of movement (*a*) to the left, (*b*) to the right, or (*c*) up or down. Combined movement reports, such as "right-up," were recorded by activation of two timers. A fourth timer automatically stopped the recording timers at the end of the 35-sec. interval. The measure used for statistical analysis was the percentage of movement reported to the right or to the left out of total movement reported, expressed as follows:

$$X = \frac{t_1 - t_2}{t_1 + t_2 + t_3} \times 100$$

where t_1 = time, in seconds, that right movement was reported; t_2 = time that left movement was reported; and t_3 = time that up-down movement was reported. Thus, a positive percentage indicates predominant movement reported to the right, a negative percentage, movement to the left. The data were treated by a double classification analysis of variance.

RESULTS

Experiment I.—The analysis of variance is summarized in Table 1. The *F* test for the interaction between age and paralysis independent of body tilt and starting position is significant. The mean positions of

TABLE 1

ANALYSIS OF VARIANCE OF EFFECT OF BODY TILT AND STARTING POSITION ON THE POSITION OF THE APPARENT VERTICAL

Source of Variance	*df*		*MS*	*F*
Between *Ss* (*S*)	39			
Between ages (A)		1	35.00[a]	1.62
Between paralyses (P)		1	4.00[a]	0.19
A × P		1	109.00[a]	5.05*
Pooled *Ss* × (A × P)		36	21.58	
Within *Ss*	440			
Between body tilt & starting position	11			
Between body tilt (B)		2	565.00[b]	26.43**
Between starting position (S)		3	1307.00[c]	88.73**
B × S		6	80.50[d]	10.45**
S × B × S	429			
B × A		2	4.00[b]	0.19
B × P		2	54.50[b]	2.55
S × A		3	97.33[c]	6.61**
S × P		3	31.67[c]	2.15
B × S × A		6	2.17[d]	0.28
B × S × P		6	13.33[d]	1.73
B × A × P		2	57.50[b]	2.69
S × A × P		3	4.00[c]	0.27
B × S × A × P		6	12.33[d]	1.60
Pooled *Ss* × (B × S)	396			
Combinations				
Pooled *Ss* × B		72	21.38	
Pooled *Ss* × S		108	14.73	
Pooled *Ss* × (B × S)		216	7.70	
Total	479			

Note.—Degrees corrected for sign.

[a] Tested against Pooled *Ss* × (A × P).

[b] Tested against Pooled *Ss* × B.

[c] Tested against Pooled *Ss* × S.

[d] Tested against Pooled *Ss* × (B × S).

* $P < .05$.

** $P < .01$.

TABLE 2

Mean Positions of Apparent Vertical (in Degrees) as a Function of Body Tilt, Starting Positions, and *S*s

Degrees	Body Tilt	Starting Position		
		Children	Adolescents	All *S*s
30° L	+2.0	−4.1	−2.8	−3.4
10° L		−2.4	−1.8	−2.1
0°	−0.4			
10° R		+2.3	+1.0	+1.8
30° R	−1.7	+4.9	+2.4	+3.6

Note.—A plus indicates a position to *S*s' right, a minus sign to the left of the objective vertical.

apparent vertical for children with right and left paralysis were +0.6 and −0.1, respectively. The comparable means for right and left paralyzed adolescents were −0.9 and +0.3. Between the children, the systematic deviation of apparent vertical is to the right in right paralytics, and to the left in left paralytics; between the adolescents, on the other hand, apparent vertical is to the left in right paralytics and to the right in left paralytics.

The *F* tests for body tilt, starting position, and starting position by age are also significant. The mean positions for each of these conditions are presented in Table 2. With respect to body tilt, apparent vertical is located relatively opposite to the side of tilt; that is, with body tilt right the apparent vertical is located relatively left, and vice versa. The means for starting position show that apparent vertical is located relatively close to the position at which the rod was started. That this effect is greater for children than adolescents is evident in the means for starting position by age.

Experiment II.—Table 3 presents a summary of the analysis of variance of the percentages after correction for sign. The *F* test for paralysis is significant, such that right paralytics, regardless of age, report predominant movement to the left (mean = −29.0%), whereas left paralytics report predominant movement relatively to the right (mean = +2.6).

Discussion

This study demonstrates that visual perception of space as reflected in the experience of verticality and apparent movement varies with enduring asymmetric muscular imbalance, such that right paralyzed children and adolescents locate apparent vertical and perceive apparent motion differently from left paralyzed children and adolescents. Furthermore, with respect to verticality, a developmental shift between children and adolescents is evident. These results further confirm the assumption of sensory-tonic field theory with regard to the importance of muscular involvement in perceptual experience. The present findings are also consistent with those of previous experiments which show that extraneous stimulation influences perception (Carini, 1955; Goldman, 1953; Wapner, Werner, & Chandler, 1951; Wapner, Werner, & Morant, 1951; Werner et al., 1951), and indicate that unilateral paralysis may be thought of as one form of extraneous stimulation. The developmental shift between children and adolescents in the experiment on apparent verticality is congruent with developmental studies of perception in

TABLE 3

Analysis of Variance of Direction of Apparent Motion in Paralyzed Children and Adolescents

Source of Variance	*df*		*F*
Between ages (A)		1	1.00
Between paralyses (P)		1	5.36*
A × P		1	0.14
Within *S*s *MS*		36	(1864.13)
Total	39		

* $P < .05$.

normal *S*s reported by Wapner and Werner (1957). That no developmental shift occurred in the experiment on apparent movement may relate to the fact that the autokinetic situation, being minimally dimensional, does not permit the subject-object differentiation assumed to be necessary for such a shift to occur.

SUMMARY

Two experiments were carried out to investigate the influence of asymmetric muscular imbalance in children and adolescents on the location of apparent vertical and on the laterality of apparent movement. Four groups of 10 *S*s each were tested. Two groups consisted of children, one with right leg paralysis, the other with left leg paralysis. The remaining two groups were composed of adolescents, one with right leg paralysis, the other with left leg paralysis.

Experiment I showed that adolescent paralytics locate the apparent vertical opposite to the side of paralysis, whereas paralytic children locate the apparent vertical to the side of paralysis. Experiment II showed that paralyzed children and adolescents report apparent movement predominantly away from the side of paralysis. These results, viewed within the framework of sensory-tonic field theory, support the assumption that muscular involvement affects perceptual experience.

REFERENCES

CARINI, L. P. An experimental investigation of perceptual behavior in schizophrenia. Unpublished doctoral dissertation, Clark University, 1955.

GOLDMAN, A. E. Studies in vicariousness: Degree of motor activity and the autokinetic phenomenon. *Amer. J. Psychol.*, 1953, **66**, 613–617.

WAPNER, S., & WERNER, H. *Perceptual development.* Worcester: Clark Univer. Press, 1957.

WAPNER, S., WERNER, H., & CHANDLER, K. A. Experiments on sensory-tonic field theory of perception: I. Effect of extraneous stimulation on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 341–345.

WAPNER, S., WERNER, H., & MORANT, R. B. Experiments on sensory-tonic field theory of perception: III. Effect of body rotation on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 351–357.

WERNER, H. *Comparative psychology of mental development.* Chicago: Follett, 1948.

WERNER, H., & WAPNER, S. Sensory-tonic field theory of perception. *J. Pers.*, 1949, **18**, 88–107.

WERNER, H., & WAPNER, S. Toward a general theory of perception. *Psychol. Rev.*, 1952, **59**, 324–338.

WERNER, H., WAPNER, S., & CHANDLER, K. A. Experiments on sensory-tonic field theory of perception: II. Effect of supported and unsupported tilt of the body on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 346–350.

(Received December 6, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 248–253

PERFORMANCE IN INSTRUMENTAL CONDITIONING AS A JOINT FUNCTION OF TIME OF DEPRIVATION AND SUCROSE CONCENTRATION[1]

JOHN R. STABLER[2]

University of Texas

As Pubols (1960) has pointed out in a recent review of the literature, there have been few experiments reported dealing with the interaction of incentive-magnitude with other variables. In particular, more research is needed in which both drive (D) and reward (K) are varied at the same time before a reasonably complete specification of the laws relating D and K can be formulated.

Ramond (1954) compared the performance of white rats in a simple runway as a function of variation in time of deprivation (T_d) and delay of reinforcement (T_g). Experiments by Fredenburg (1956) and Besch and Reynolds (1958) involved the simultaneous manipulation of T_d and length of the response chain (L_g), i.e., distance that S had to run to reach the goal. The data from these experiments indicated no interaction between D and K; however manipulating either T_g or L_g introduces variable opportunity for competing responses to occur. A fatigue variable may also be differentially present when L_g is varied. The results from studies which have manipulated the amount of the incentive (W_G) are contradictory (Pubols, 1960). One reasonable explanation for the inconclusiveness of the results from such studies is suggested by Guttman's (1953) analysis of this particular method of varying reward magnitude. He points out that a variation in the size or number of the incentive object introduces several simultaneous sources of variation. A large incentive object has different visual, olfactory, and tactile stimulus features and requires more chewing and swallowing than a small incentive object. The purpose of the present study was to determine the nature of the relationship between D and K when different sucrose concentrations are used to vary K, thereby controlling for the possibility of differences in preconsummatory stimulation, ingestive activity, and presence of competing responses.

A second purpose of the present experiment was to provide additional information on the dual reinforcing effect of sucrose solutions. Previous experimentation (Wetzel, 1959) has indicated that rats inexperienced with sucrose respond to it on the basis of its palatibility. On the other hand, if hungry *S*s are fed adequate quantities of sucrose solutions over a period of days, they will subsequently respond to the solution not only for the sensory stimulation that it provides but also for its hunger-reducing potential. By employing a within-groups instead of a between-groups design, the present experiment provides evidence as to whether or not the second reinforcing property of sucrose becomes manifest over a prolonged period of small-quantity feedings.

---

[1] Part of a dissertation submitted to the Graduate School of the University of Texas in partial fulfillment of the requirements for the PhD degree. The author is indebted to Alfred Castaneda for advise and assistance throughout the course of the investigation.

[2] Now at Louisiana State University.

METHOD

Subjects and apparatus.—The *S*s were 90 experimentally naive male rats of the Wistar strain purchased from Harlan Industries, Cumberland, Indiana. All *S*s were 85–90 days old when purchased and 106-111 days old at the beginning of pretraining.

The apparatus was a straight runway consisting of a 12-in. start box, a 24-in. alley, and a 12-in. goal box. The runway was 6 in. high and 3 in. wide. It was covered by wire mesh. All inside spaces were painted flat black. A 100-w. bulb was located 55 in. above each end of the alley. The light from these bulbs was shaded by a 60 × 12 in. sheet of fiberboard. The start box and alley were separated by two guillotine doors, one of clear plastic and the other of opaque fiberboard.

Running-time measures were provided by interruption of monochromatic red beams to photocells which controlled relays and standard timers. The raising of the plastic door started the first clock. Interruption of the first beam, 6 in. into the alley, stopped the first clock and started a second. Interruption of a second beam, located 30 in. into the alley, stopped the second clock and started a third, which stopped when a beam located 11 in. into the goal box was interrupted. The final beam was located directly over the food cup.

Experimental design.—In Stage 1 of training, two levels of hunger (4 hr. and 20 hr.) and three levels of sucrose concentration (8%, 16%, and 32%) were employed in a 2 × 3 factorial design. The *S*s were randomly assigned to the six experimental conditions designated as: D_4K_8, D_4K_{16}, D_4K_{32}, $D_{20}K_8$, $D_{20}K_{16}$, and $D_{20}K_{32}$. In each of two groups 1 *S* died, and in a third group, 2 *S*s died. To equate the number of *S*s within groups, 1 randomly selected *S* was discarded from each of the three remaining groups, and the missing data for 1 *S*, which died after the completion of Stage 1, were prorated from the scores of *S*s remaining within its group. This left 14 *S*s in each group, 84 *S*s in all. The three sucrose concentrations were designated according to percentage of concentration by weight. Thus, the 8% solution was prepared by combining 8 gm. of sucrose (commercial table sugar) and 92 cc of distilled water. The solutions were kept in closed containers and were prepared fresh before each day of testing.

In Stage 2, each group was divided into matched experimental and control groups of 7 *S*s each. This division resulted in a 2 × 2 × 3 factorial design. Fifteen minutes after each testing session the experimental *S*s were allowed 15 min. access to sugar solution of the same percentage as that concentration whicht hey had consumed in the goal box. The control *S*s continued under the same conditions as in the first stage. All *S*s were fed 2 hr. after the time of measurement.

Procedure.—Two weeks prior to the first training day, *S*s were placed on their respective food deprivation schedules. Water was continuously available except during the experimental sessions. During the 10 days prior to the first training day, *S*s were adapted to handling and to the apparatus. The adaptation procedure closely approximated the procedure to be followed during the experiment.

Each *S* was given three trials a day on each of the 20 days of Stage 1 and each of the 20 days of Stage 2. The following procedures were employed on every trial for each *S*: *S* was placed in the start box, and the opaque door was opened as soon as *S* oriented to it. The plastic door was removed ½ sec. later. As soon as the two drops of incentive substance had been consumed, *S* was removed and placed in a waiting cage until the next *S* had run a trial, when it was returned to its home cage for the remainder of the intertrial interval. Intertrial intervals were approximately 20 min. long. The number of times that *S* stopped its forward movement in each alley segment and the time taken to traverse each segment were recorded after every trial.

RESULTS

Running speed.—The final performance levels attained under the six experimental conditions were of primary interest; hence the mean running speeds (reciprocal time in ft/sec) over Trials 31–60 were computed for each *S*. The group mean running speeds in the start box and in the runway for Trials 31–60 are plotted in Fig. 1 and 2. Because the results of the analyses of variance of the mean goal box speeds were comparable to those of the start box and runway, i.e., all showed a significant interaction, the data based on that particular segment were not further analyzed. Since previous research (Besch & Reynolds, 1958) has suggested that

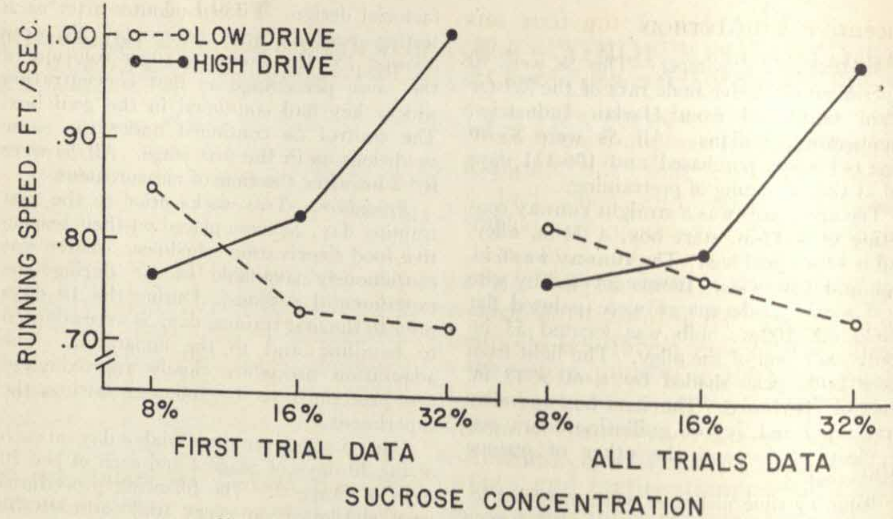


FIG 1. Mean start box speeds for Trials 31–60.

data based on the first trial of each day, rather than on all trials of each day, may differ with respect to amount of *D* and *K* interaction, both types of data were analyzed in the present study. Start box and runway measures and measures based upon either the first trial of each day or upon all trials of each day were generally consistent. There was, however, some evidence of a "warm up" effect from the first trial of each day. Runway speed measures based on all trials of each day were higher than those based on only the first trial of each day.

A significant interaction was found between the effects of drive and

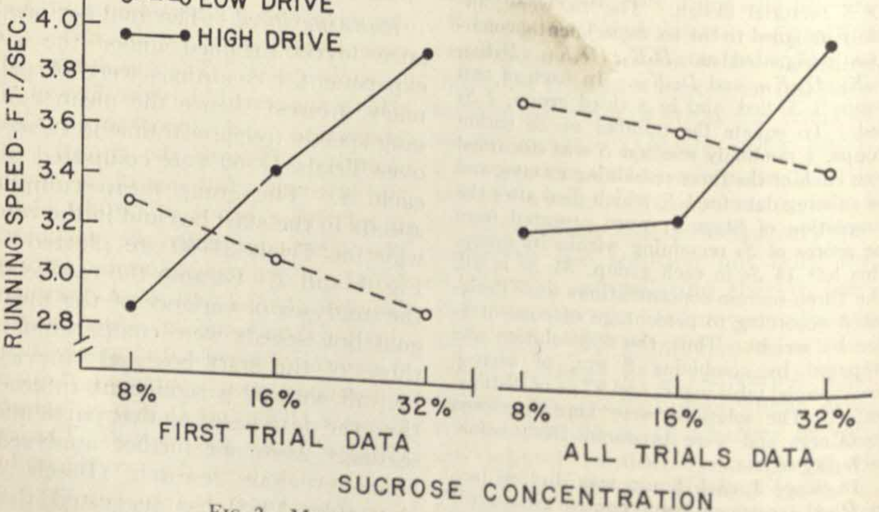


FIG. 2. Mean runway speeds for Trials 31–60.

incentive with each of the four sets of data ($F = 4.58, P < .05; F = 4.94, P < .01; F = 4.19, P < .05; F = 3.15, P < .05$; respectively, with $df = 2/78$ in each case). Trend analyses (Grant, 1956) indicated that the variance associated with interaction had a highly significant linear component (Fs = 9.16, 9.21, 8.38, and 6.11, respectively, with $P < .01$, $df = 1/78$ in all cases). Duncan multiple range tests (MRT) were performed in order to make orthogonal comparisons between the six experimental groups. The four tests consistently showed that Group $D_{20}K_{32}$ was significantly faster than Group $D_{20}K_8$. This difference was the only significant one obtained from the mean runway speeds based on all trials of each day. According to tests of start box and runway speeds based on the first trial of each day, Group $D_{20}K_{32}$ was also significantly faster than either Group D_4K_{16} or Group D_4K_{32}. The MRT performed on the start box speeds based on all trials of each day showed a significant difference between Group $D_{20}K_{32}$ and each of the others. Although the graphs indicate an inversion of drive groups at K_8, none of the MRTs showed a significant difference between Groups D_4K_8 and $D_{20}K_8$. It should also be noted that in no case were the means of Groups D_4K_8 and D_4K_{32} significantly different from each other.

TABLE 1

NUMBER OF STOPS FOR TRIALS 31–60, BASED ON ALL TRIALS FOR EACH DAY

Group	Alley Segment					
	Start Box		Runway		Goal Box	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
D_4K_8	3.1	3.13	2.8	2.88	3.1	3.84
D_4K_{16}	4.9	4.02	4.0	2.95	3.1	3.55
D_4K_{32}	5.9	3.50	2.1	1.35	1.4	2.09
$D_{20}K_8$	5.3	4.78	3.7	2.94	3.1	3.57
$D_{20}K_{16}$	4.3	4.48	2.1	2.22	1.5	1.63
$D_{20}K_{32}$	1.4	1.84	.7	1.22	.5	1.05

TABLE 2

MEAN RUNNING SPEEDS FOR TRIALS 91–120, BASED ON ALL TRIALS FOR EACH DAY

Group	Alley Segment			
	Start Box		Runway	
	Exp.	Control	Exp.	Control
D_4K_8	.94	1.00	4.03	3.76
D_4K_{16}	.89	.86	3.78	3.32
D_4K_{32}	.88	.95	3.26	3.37
$D_{20}K_8$	.96	1.04	3.72	3.67
$D_{20}K_{16}$	1.08	.92	3.84	3.74
$D_{20}K_{32}$	1.11	1.18	4.19	4.46

Stop behavior.—The mean numbers of stops for each group in the start box, the runway, and the goal box for Trials 31–60, based on all trials of each day, are presented in Table 1. A significant interaction was obtained from the analysis of variance of the number of stops in the start box ($F = 5.03$; $P < .01$, $df = 2/78$); however the runway and the goal box analyses showed only a significant incentive effect (Fs = 4.06 and 3.79, respectively, with $P < .05$, $df = 2/78$ in each case). The results of the MRTs suggested that the loss of the interaction and appearance of an incentive effect in the runway and goal box was a function of the change in relative position of Group D_4K_{32}. The significant difference which was found in running speeds between Groups $D_{20}K_{32}$ and $D_{20}K_8$ occurred also for stop behavior. Also, as with the running speed data, no significant differences were obtained between Groups D_4K_8 and $D_{20}K_8$, or between Groups D_4K_8 and D_4K_{32}. The correlations between the two dependent variables for start box, runway, and goal box, based on all trials of each day, were −.72, −.63, and −.42, respectively.

Results of Stage 2.—The mean running speeds for each group in the start box and the runway for Trials 91–120, based on all trials of each day, are presented in Table 2. Analyses of the data showed a reliable D and K interaction effect for runway speeds ($F = 5.66$, $P < .01$, $df = 2/72$) and a reliable D effect for both start box and runway speeds ($F = 13.09$, $P < .01$; $F = 5.38$, $P < .05$; respectively, with $df = 1/72$ in each case). It may be recalled that no significant D effect was obtained from the analyses of Trials 31–60. Multiple range tests performed on start box and runway data showed no significant differences between experimental and control groups at any given D and K combination of conditions. For example, although the experimental and control groups of combination $D_{20}K_{32}$ ran significantly faster than several other groups, they did not differ significantly from each other. The groups representing the various D and K combinations held the same relative positions at the end of Stage 2 as they had held at the end of Stage 1. The start box and runway data for Trials 91–120, based on all trials of each day, were also covaried upon the individual means of Trials 31–60 in order to control statistically for the level of performance that had been attained before the beginning of Stage 2. The D effect was found to be highly reliable ($F = 10.13$, $F = 13.10$; $P < .01$, $df = 1/71$ in each case).

DISCUSSION

Relation of D *and* K *variables.*—The evidence from this experiment is clearly more compatible with Hull's position that drives and incentives are multiplicatively related than with Spence's proposal that drives and incentives join in an additive fashion. One implication of the results is that the method used to vary the amount of reinforcement, i.e., L_g, T_g, W_G, etc., is of critical importance to theoretical predictions regarding incentive effects.

With the exception of Group D_4K_{32}, the stop data appeared to be inversely related to running speed measures. This is consistent with the observations of Cotton (1953) and others that running speed measures are partly dependent upon the number of times S stops in the alley or performs some other nonadjustive response.

Hypotheses about incentive action.—Spence (1956) has proposed that one determinant of the strength of an incentive is the generalized habit strength (which here will be denoted $\bar{H}_K$) between alley stimuli and anticipatory consummatory behavior ($r_g - s_g$). Several recent eyelid conditioning experiments (Spence, Haggard, & Ross, 1958; Trapold & Spence, 1960) have indicated that the magnitude of H varies as a function of the intensity of the unconditioned stimulus (UCS). In such respects as the abruptness of the change in sensory stimulation which it produces, using a sucrose solution as the UCS appears analogous to an air-puff UCS; thus incentive magnitude in the present study may have been a function of differential strength of $\bar{H}_K$. If the equation for excitatory potential (E) were to be written $E = (H + \bar{H}_K) \times D$, it would be possible to derive that D and K combine multiplicatively as suggested by the present experiment.

$$E_{\mathrm{High}} = (H + \bar{H}_{K_{\mathrm{High}}}) \times D$$
$$E_{\mathrm{Low}} = (H + \bar{H}_{K_{\mathrm{Low}}}) \times D$$

Subtracting:

$$E_{\mathrm{High}} - E_{\mathrm{Low}} = D(\bar{H}_{K_{\mathrm{High}}} - \bar{H}_{K_{\mathrm{Low}}})$$

Dual reinforcing effect of sucrose solutions.—The finding of a significant D effect at the end of Stage 2, together with the absence of a significant D effect at the end of Stage 1, indicates that the effect of level of deprivation upon running speed is greater if *S*s have been rewarded with small amounts of sucrose solution over an extended period of time. It suggests that *S*s learned to respond to the incentive in terms of its need-reducing potential as well as in terms of its palatability. The absence of reliable experimental versus control main effects and interaction effects indicates that the *D* effect was determined by experience with sucrose rewards on the 60 trials of Stage 2 rather than by extra sucrose consumption in the home cages.

SUMMARY

The experiment was primarily concerned with the joint effects of concentration of sucrose incentive and hours of food deprivation upon locomotor behavior in a straight alley. Three levels of concentration and two levels of drive were employed in a factorial design with six groups of male rats. Running speed and frequency of stops were recorded for 120 trials, i.e., 60 trials in Stage 1 and 60 trials in Stage 2.

The major finding from Stage 1 was that asymptotic running speed measures showed a significant interaction between drive and incentive variables. Frequency of stops was generally inversely related to running speed. Start box and runway measures, and measures based upon either the first trial of each day or upon all trials of each day, were generally consistent. The results were discussed in terms of Spence's (1956) incentive motivation hypotheses concerning the action of incentives in instrumental reward conditioning.

In Stage 2 half the *S*s in each experimental group were allowed to consume large quantities of the same concentration sucrose solution as they had been receiving in the goal box. The solutions were provided after each day's testing and 90 min. before feeding. A significant *D* effect, obtained from analyses of the asymptotic running speeds at the end of the second stage, was interpreted as evidence that after a prolonged period of small-quantity feedings of sucrose solution *S*s came to respond to the incentive in terms of its need-reducing potential as well as in terms of its palatability.

REFERENCES

BESCH, N. F., & REYNOLDS, W. F. Alley length and time of food deprivation in instrumental reward learning. *J. exp. Psychol.*, 1958, **56**, 448–452.

COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

FREDENBURG, N. C. Response strength as a function of alley length and time of deprivation. *Proc. Ia. Acad. Sci.*, 1956, **63**, 563–567.

GRANT, D. A. Analysis of variance in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–155.

GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

PUBOLS, B. H. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

RAMOND, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, **47**, 248–251.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.

TRAPOLD, M. A., & SPENCE, K. W. Performance changes in eyelid conditioning as related to the motivational and reinforcing properties of UCS. *J. exp. Psychol.*, 1960, **59**, 209–214.

WETZEL, R. J. The effect of experience with a taste reward. *J. comp. physiol. Psychol.*, 1959, **52**, 267–272.

(Received December 9, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 254–257

CONCEPT IDENTIFICATION OF AUDITORY STIMULI AS A FUNCTION OF AMOUNT OF RELEVANT AND IRRELEVANT INFORMATION[1]

ROSARIA G. BULGARELLA[2] AND E. JAMES ARCHER

University of Wisconsin

A number of experiments (Archer, Bourne, & Brown, 1955; Bourne, 1957; Brown & Archer, 1956) have studied the effect of variations in the amount of nonredundant irrelevant information on concept identification. In general, a linear degradation of performance is produced by an increase in irrelevant information.

Two additional studies (Walker, 1957; Wargo, 1960) have investigated the concurrent effect of variation of nonredundant relevant information. In both of these studies, task difficulty increased as a function of an increase in relevant and irrelevant information. However, Walker found the increase to be nonlinear and also observed a significant interaction between the two variables. The failure of Wargo to find either of these effects was assumed to be due to certain metholological differences which will be considered later.

Except for a study by Lordahl (1959), all of the previous concept identification studies have used visual stimuli only. In the exception, Lordahl presented problems which required *S* to use 1 bit of relevant visual information and 1 bit of relevant auditory information while *E* simultaneously varied the amount of irrelevant information to both sense modalities. Whereas the effect of such variation was as expected when the irrelevant information was visual, the auditory irrelevant information did not have a significant adverse effect on performance.

The purpose of the present study was to determine whether, in general, auditory information was comparable to visual information in concept identification, and in particular to determine the effect of simultaneous variation of both relevant and irrelevant *auditory* information.

PROCEDURE

Subjects.—Ninety women students from introductory and intermediate courses in psychology at the University of Wisconsin served individually for one session each. All were volunteers who had not previously served in a concept identification study.

Design.—The major independent variables were the amounts of relevant and irrelevant auditory information. A 3 × 3 orthogonal design with 1, 2, or 3 bits of relevant information and 0, 1, or 2 bits of irrelevant information was used.

Five different specific combinations of relevant information defined the problems at each level of relevant information, giving 15 problems in all. Since the same problems were presented at each of the three levels of irrelevant information, there were 45 different combinations of problems and irrelevant information. Two *S*s were randomly assigned to each of these 45 combinations.

Apparatus.—The auditory stimuli were presented through a pair of earphones worn by *S* who sat before a control console. The console had two lever-action switches and two lamps. Above one switch a black plus sign was painted. This switch was to be pressed for all positive instances of the concept. The other switch was unmarked, and it was to be pressed for all negative instances of the concept. When *S* made a correct response, a white lamp came *on*; when *S* made an incorrect response, a red lamp came *on*. In both cases the feedback signal continued until the next auditory pattern was presented 4 sec. later.

---

[1] This research was supported by a grant from the Research Committe from funds provided by the Wisconsin Alumni Research Foundation. The paper was prepared while the second-named author was a Senior Postdoctoral Fellow (NSF) at the Applied Psychology Research Unit, Cambridge, England.

[2] Now at Michigan State University.

The apparatus which generated the auditory stimuli has been described in a previous study by Lordahl (1959). The *S* saw only the control console; the remainder of the apparatus was concealed by a partition. After giving the instructions, *E* sat behind the partition and was not visible to *S*. An Esterline-Angus 20-channel operations recorder was used to obtain a permanent record of *S*'s responses. A counter recorded *S*'s errors, and a stopwatch was used to measure the time to criterion.

Instructions and task.—The instructions were tape recorded and played to *S* at the beginning of the session. The *S* was informed that she would hear various sounds which were to be placed in the categories represented by the switches before her. She was told that those sounds which were to be placed in the category represented by the switch with a plus sign had something in common with each other, and all other sounds belonged in the unmarked category. The significance of the feedback lamps was explained and *S* was instructed to press only one switch for each tone pattern. Accuracy was stressed over speed.

The task was self-paced (*S* could take as long as she liked to respond to a tone pattern), but the next stimulus was automatically presented 4 sec. after the response was made. The *S* served until she had made either 32 consecutively correct responses or had heard a total of 192 stimuli, whichever came first. If *S* did not reach the performance criterion in 96 patterns, she was given a 1-min. rest before the session was continued.

Stimuli and problems.—The auditory stimuli could be described in terms of five bileveled dimensions. The *frequency* of the tone was either 250 cps or 1800 cps. Its *intensity* was either 43 or 64 phons for the 250-cps tone, or it was 54 or 64 phons for the 1800-cps tone. The intermittent addition of *white noise* (at 55 phons) or its absence was another dimension. The sound also had a quality of *steadiness* or regular interruption (on for 1 sec. and off for .5 sec.). The fifth dimension was the *earphone*, either left or right, through which the sound was presented.

When variation of a dimension was not needed, as in some problems where the dimension was neither relevant nor irrelevant, one level was consistently presented. The constant levels were 1800 cps for frequency, 43 or 54 phons (depending on which of the frequencies was used) for the intensity, the absence of white noise, steadiness of the tone, and the left earphone. The specific levels of the dimensions were determined by a pilot study which required comparable *S*s to show 100% discriminability of the levels.[3]

The 15 problems were of three types. Five problems had only 1 bit of information relevant. For example, when frequency was the only relevant information *S* would hear soft, clean, steady tones in the left earphone. If the tone had a high frequency, *S* was to press the switch with the plus sign. If the tone had a low frequency, *S* was to press the switch which was unmarked.

Five other problems had 2 bits of relevant information. For example, if frequency and intensity were the relevant dimensions, *S* would hear a clean, steady tone in the left earphone. If the tone had both a high frequency and a low intensity, *S* was to press the switch with the plus sign. For all other tones *S* was to press the unmarked switch.

In both of these examples no irrelevant information was present. If, however, earphone had been an irrelevant dimension, then although the above described tones would have come into each ear equally often, *S* should ignore this dimension and respond only to frequency and intensity in the latter problem, and only frequency in the former.

The third set of five problems had 3 bits of relevant information. For example, a positive instance could be a high frequency, low intensity, and interrupted tone. All others would be negative instances.

The program, which was controlled by Western Union tape transmitters, contained a series of 96 stimuli. If the concept was not identified in the first presentation of 96 patterns, the series was repeated.

The order of presentation of the negative and positive instances of the concept was determined randomly, but no matter how much information was relevant or irrelevant, 25% of the presented patterns were positive instances.

RESULTS

Trials to criterion.—As the amount of relevant information increased, the number of trials (auditory patterns) increased. For 1, 2, and 3 bits of

---

[3] The writers are indebted to Willard R. Thurlow for his help in assessing the frequencies and intensities of the signals.

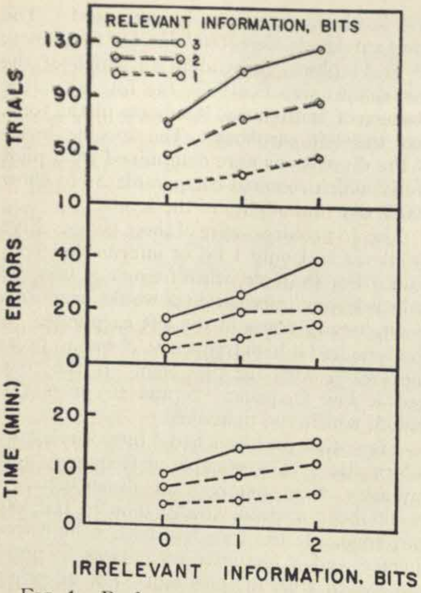


FIG. 1. Performance as measured by mean number of trials, mean number of errors, and mean time to criterion as a function of amount of nonredundant relevant and irrelevant information.

relevant information, the mean number of trials to criterion was 62.6, 99.3, and 131.4, respectively. The few *S*s who failed to achieve the performance criterion were assigned the score of 192. Failing to take all *S*s to the performance criterion introduces a slight bias in estimate for high levels of both relevant and irrelevant information. On the other hand, replacement of these less able *S*s with others who could solve the problem would have introduced a bias in the opposite direction. Figure 1 shows the mean number of trials to criterion as a function of both relevant and irrelevant information. The results of an analysis of variance indicated that the increase in trials to criterion was highly significant ($F = 23.48$, $df = 2/12$, $P < .001$) and that this function is best described by a straight line since only the linear component was significant.

For 0, 1, and 2 bits of irrelevant information, *S*s required 74.8, 103.1, and 112.0 trials, respectively, to achieve the criterion. This increase was significant ($F = 9.18$, $df = 2/45$, $P < .001$). In this case too, only the linear component was significant; the increasing performance degradation is well described by a straight line.

Of particular interest is the lack of an interaction between relevant and irrelevant information. At least over the range studied, adding irrelevant information into an already difficult auditory display did not make the task differentially more difficult.

Errors.—This measure is the number of stimuli incorrectly assigned to a category by *S*. Figure 1 shows the mean number of errors as a function of relevant and irrelevant information. For 1, 2, and 3 bits of relevant information, the mean number of errors was 9.3, 17.5, and 27.6, respectively. These differences were highly significant ($F = 25.70$, $df = 2/12$, $P < .001$). The function was best described by a straight line since only the linear component was significant.

A similar increase in number of errors was observed as a function of irrelevant information: the mean number of errors was 9.7, 18.8, and 25.8 for 0, 1, and 2 bits of irrelevant information, respectively. These differences were also highly significant ($F = 8.24$, $df = 2/45$, $P < .001$), and again only the linear component was significant.

Time to criterion.—Figure 1 shows the time to criterion as a function of both relevant and irrelevant information. The results of an analysis of variance were similar to those for trials to criterion and errors. Both relevant and irrelevant information and their linear components were significant and none of the other terms was significant.

DISCUSSION

The results of the present study indicate that concept identification of auditory stimuli is quite comparable to concept identification of visual stimuli. As in previous studies (Bourne, 1957; Brown & Archer, 1956; Walker, 1957; Wargo, 1960) performance, as measured by trials, errors, and time to criterion, degraded as a linear function of amount of nonredundant irrelevant information. Only in the study by Archer, Bourne, and Brown (1955) was the performance degradation nonlinear. Apparently that finding was anomalous.

The present results are also consistent with those of Wargo (1960). Performance degraded as a linear function of nonredundant relevant information and the interaction between relevant and irrelevant information was not significant. In this regard the present study is also not consistent with Walker (1957). Since Walker required his *S*s to use two, four, or eight response categories when 1, 2, or 3 bits of information were relevant, response complexity was confounded with concept complexity. The differential response complexity could account for the nonlinear performance degradation he observed; and this nonlinear function would, in turn, account for the significant interaction.

Of particular interest is the contrast of Lordahl's (1959) data with the present results. In his study, variation of irrelevant auditory information had little effect on performance when *S* had to use both visual and auditory relevant information simultaneously. Apparently, as Lordahl suggested, *S*s either have a preference for responding to visual irrelevant information or they are better able to ignore auditory irrelevant information. However, when the concept to be identified involves only auditory signals, then variation in auditory irrelevant information is just as effective a way of manipulating the complexity of the task.

SUMMARY

The effect upon concept identification of the simultaneous variation of nonredundant relevant and irrelevant auditory information was determined. The design had three levels (1, 2, or 3 bits) of relevant information orthogonal to three levels (0, 1, or 2 bits) of irrelevant information. There were five different problems within each of the three levels of relevant information. Two *S*s were assigned to each of the 45 combinations of problems and information.

The *S*'s task was to place auditory stimuli into one of two categories, i.e., distinguish between positive and negative instances of conjunctive concepts. Ninety *S*s served individually to a criterion of 32 consecutively correct responses or a maximum of 192 trials (whichever occurred first). The *S*s received immediate information feedback about the correctness of their responses through two signal lights.

The major results were: (*a*) Performance, as measured by trials, errors, and time to criterion, degraded as a linear function of both nonredundant relevant and irrelevant information. (*b*) The interaction between the two variables of relevant and irrelevant information was not significant.

REFERENCES

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

LORDAHL, D. S. Concept identification as a function of the amount of information in simultaneously presented visual and auditory signals. Unpublished doctoral dissertation, University of Wisconsin, 1959.

WALKER, C. M. Concept identification as a function of amounts of relevant and irrelevant information. Unpublished doctoral dissertation, University of Utah, 1957.

WARGO, D. G. Concept identification as a function of relevant and irrelevant information. Unpublished master's thesis, University of Wisconsin, 1960.

(Received December 9, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 258–268

TRANSFER FROM VERBAL-DISCRIMINATION TO PAIRED-ASSOCIATE LEARNING[1]

WILLIAM F. BATTIG, JOHN M. WILLIAMS,[2]
AND JOHN G. WILLIAMS[3]
University of Virginia

The need for a more analytic approach to verbal learning, whereby two or more phases or stages are distinguished and treated separately, is becoming increasingly apparent. Underwood and his associates have clearly demonstrated the importance in verbal paired-associate learning of distinguishing between a *response-learning* or *response-recall* phase and an *associative* phase (Underwood, Runquist, & Schulz, 1959), and have further suggested that quite different mechanisms may be involved in these two phases (Underwood & Schulz, 1960). The implications of a similar stage-wise analysis have also been profitably exploited in studies of serial learning (Hovland & Kurtz, 1952), as well as verbal-discrimination learning (Runquist & Freeman, 1960).

In view of the high order of complexity of verbal paired-associate learning, an extension of this kind of analytic approach to encompass more than two phases would seem to be a strong possibility, at least at a conceptual level. However, the usefulness of the Underwood et al. approach can be largely attributed to their success in separating response-learning and associative phases in terms of specific experimental operations involving some type of familiarization with the material to be learned prior to its use in a paired-associate task. Further analyses of paired-associate learning into separate stages will therefore be of value only to the extent that experimental procedures can be devised which permit the isolation and separate evaluation of these phases.

The process of discriminating between items of the list has been suggested previously as an important aspect of paired-associate learning (e.g., Gibson, 1940), and may represent an additional phase separable from either the response-learning or associative phases. Appropriate operations for the experimental isolation of a discrimination phase would seem to be provided through the use of a verbal-discrimination (VD) learning procedure prior to paired-associate (PA) learning. Accordingly, the first of the present experiments represented an attempt to isolate and determine the relative importance of a phase in PA learning involving the discrimination between items of a pair, as compared with familiarization and associative phases similar to those identified previously. This was accomplished by introducing controlled variations in the relationships existing between the material used in VD learning, and that subsequently employed in PA learning. Due to the unexpected nature of the results, which gave no evidence of any positive effect on PA performance from either prior discrimination or familiarization by means of a VD task, Exp. II was then undertaken to try to replicate and extend these findings concerning the nature of transfer between VD and PA learning.

---

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

[2] Now at College of William and Mary.

[3] Now at Carnegie Institute of Technology.

METHOD

Experiment I

Subjects.—The 36 *S*s were undergraduates who were either paid for their services or satisfied an introductory psychology course requirement. Additional motivation was provided by an announced bonus of $5 and $2 to the two fastest learners. Each *S* served for a single session of 1 hr. or less, during which he first learned a VD list, immediately followed by a PA list.

Materials.—The nonsense-syllable materials were all selected from the 34–37% range of association values as determined by Archer (1960). A single list of 12 pairs was constructed for PA learning, taking special precautions to minimize formal similarity between any of the items in the list and to avoid obvious associations within any pair. An additional eight syllables, chosen so as to be minimally similar to those on the PA list and to one another, were selected for use in the VD task, in conjunction with the syllables that were also used in the PA list.

VD procedure.—Each *S* was presented with a list of 12 pairs of nonsense-syllables, one of each pair having been previously designated as "right" and the other "wrong," and instructed to learn to respond by pronouncing the right syllable each time that pair was presented. Each pair was typed on a 3 × 5 in. card and presented by *E* through a manually operated card exposure device under self-paced conditions, until *S* had pronounced one of the two syllables on the card. At this point, *E* said either "right" or "wrong," after which the card containing the next pair was presented immediately. The use of self-paced conditions was necessary to insure that a response was made to each pair, although *S*s typically responded within 2 or 3 sec., and few deviated appreciably from a regular rate of responding. A prearranged order of presentation provided that *(a)* all 12 pairs were presented once on each trial; *(b)* no pair appeared in the same position within the list on more than one trial (unless *S* required more than 12 trials, in which case the orders used on the first 12 trials were repeated); *(c)* each syllable appeared in the left and right positions on the card equally often; *(d)* an approximately equal number of right and wrong syllables appeared in each position on the card within any trial. Learning was carried out to a criterion of two successive errorless trials. Immediately following the conclusion of VD learning, a recall test was given in which the wrong syllable of each pair was presented individually while *S* attempted to recall the right syllable previously paired with it, in order to check on the extent to which intrapair associations had been learned during the VD task. No information was given as to the correctness of these recall responses, in order to prevent the actual learning of associations during the recall test.

PA procedure.—Having been given no previous indication that anything beyond the VD task would be required of him, *S* was taken immediately after the conclusion of VD learning to another part of the experimental room previously blocked from his view, and seated before the memory drum on which the PA task was presented. Standard instructions for PA anticipation learning were given, with an added statement that some of the syllables had been used previously on the VD task, so that *S* should try to anticipate the correct response the first time through the list. The PA learning task was identical for all *S*s, with each stimulus being presented alone for 2 sec., while *S* attempted to pronounce the correct response, followed by the stimulus-response pair for 2 sec., with an 8-sec. intertrial interval. Each *S* learned to a criterion of 2 successive errorless trials, or to a maximum of 24 trials.

Design.—A mixed-lists within-*S*s design was employed, whereby the 12 pairs were systematically varied in terms of their relationship to the previous VD task, as follows:

Discriminated (D) pairs: Four of the 12 PA pairs had also been used as pairs in the VD list. For 2 of these pairs, representing the D(S) condition, the PA stimulus had been "right" and the response "wrong" on the VD task. For the other 2 pairs, the D(R) condition, the response had been right and the stimulus wrong. As defined for purposes of this study, these pairs were assumed to have gone through both familiarization and discrimination phases during the VD task, but not the associative phase.

Familiarized (F) pairs: An additional four PA pairs consisted of syllables previously used in the VD list, but each had been paired therein with another syllable not used in the PA task. As with the D pairs, two of these F pairs consisted of stimuli which were right and responses wrong on the VD task, while the reverse was true for the other two pairs, which are designated as the F(S) and F(R) conditions, respectively. The familiarization phase was therefore assumed to have already been provided by the VD task for these pairs, but they had not been directly discriminated,

TABLE 1

VD LISTS ILLUSTRATING CONDITIONS FOR EXP. I AND II, AND PA LIST USED FOR ALL *S*S OF BOTH EXPERIMENTS

Experiment I		PA List (Exp. I & II)	Experiment II	
VD List	Cond.		Cond.	VD List
R[a] W		S R[b]		R[a] W
BIY-ROJ	D(S)	BIY-ROJ	D(S)	BIY-ROJ
JAF-LIH	D(S)	JAF-LIH	D(S)	JAF-LIH
KUX-VOR	D(R)	VOR-KUX	D(R)	KUX-VOR
PIB-QAW	D(R)	QAW-PIB	D(R)	PIB-QAW
RUZ-ciq	F(S)	RUZ-TEY	L(S-R)	RUZ-ZOG
yun-TEY	F(S)	ZOG-DEQ	L	TEY-VAK
ZOG-qer	F(R)	PEH-VAK	L(S)	PEH-SUQ
gah-DEQ	F(R)	SUQ-GED	L(R)	GED-DEQ
bov-PEH	N	TIV-JEC	U(S-R)	TIV-QUM
VAK-zes	N	NUK-ZIL	U	JEC-YIS
jiz-SUQ	N	WOX-QUM	U(S)	WOX-ZIL
GED-jub	N	YIS-MUP	U(R)	MUP-NUK

Note.—VD syllables not used in PA list written in lower-case letters.

[a] R = right; W = wrong.

[b] S = stimulus; R = response.

either from the syllables paired with them or from any other syllable used in the PA list.

New (*N*) *pairs*: The final four PA pairs consisted of syllables which had not been used in the VD list, so that neither prior discrimination nor familiarization phases had taken place.

One of the six VD lists is presented in the left column of Table 1, to illustrate the method by which each condition was produced within the framework of a single PA list. By means of a latin square, each subset of four PA pairs was used under the D, F, and N conditions for one-third of the *S*s. In addition, the right and wrong syllables of each VD pair were reversed for half of the *S*s within each of these three subgroups, so that each PA pair also appeared equally often within the S and R subsets of the D and F conditions. As a consequence, a total of six different VD lists were used, each being learned by 6 *S*s. Each of the first 6 *S*s to appear in the laboratory was assigned to a different VD list on the basis of a predetermined random order, as was also true for each subsequent set of 6 *S*s.

Experiment II

In order to check the possibility that the results of Exp. I were due to its within-*S*s design, so that performance on the N condition of the PA task may have been facilitated by the relative novelty of these pairs, a separate control group of *S*s was tested in Exp. II under conditions corresponding to the N condition of Exp. I. In addition, PA pairs consisting of syllables previously paired on the VD task with syllables from other pairs of the PA list were employed, replacing the F condition of Exp. I. Furthermore, besides the variation in the PA pairs of the stimulus or response position of the right and wrong syllables from the VD task, the number of right or wrong syllables within the PA pairs was also systematically varied. These and other minor differences from Exp. I are described more fully below, although in most respects the materials and procedure were identical for the two experiments.

Subjects.—The *S*s were 60 summer session students who were paid for their services, and were assigned to an experimental group of 36 or a control group of 24 *S*s in terms of a predetermined random order based on their order of appearance in the laboratory. An additional 8 *S*s were eliminated due to their inability to learn the VD list to criterion within a reasonable time, and were replaced by the next *S* to be tested.

Materials.—The same nonsense syllables used in Exp. I were employed, along with an additional 16 syllables selected from the 32–41% range of association values (Archer, 1960) for the VD task of the control group.

Design.—Prior to PA learning, the control group learned a VD list consisting exclusively of syllables not used in the PA list. For the experimental group, all syllables of the PA task had also been used in the VD task, but the two tasks were related in several different ways. Four of the 12 PA pairs were identical to the D(S) and D(R) conditions of Exp. I. Four other PA pairs constituted a like-position (L) condition wherein each stimulus had been paired on the VD task with a stimulus from another PA pair, and each response with another response. Each of these pairs differed in the number or position of syllables which were right or wrong on the VD task, as follows: (*a*) L(S-R) condition: Both syllables of the PA pair had been right on the VD task; (*b*) L condition: Neither syllable was right on the VD task; (*c*) L(S) condition: The stimulus syllable was right and the response wrong; (*d*) L(R) condition: The response was right and the stimulus wrong. The final four pairs represented an unlike-position (U) condition, in that each PA stimulus had been paired on the VD task with a response from another PA pair, and each response with the stimulus from another pair.

As with the L conditions, these represented all possible combinations of right and wrong syllables, and were also subdivided into U(S-R), U, U(S), and U(R) pairs.

As in Exp. I, a latin square counterbalancing procedure was employed so that each pair of the PA list was used equally often under each condition. Due to the increase in number of conditions, the use of 12 different VD lists was necessary to accomplish this, each being learned by 3 *S*s of the experimental group. One of these lists is presented in the right half of Table 1, to illustrate the method by which the various relationships between the VD and PA pairs were produced.

Other minor differences from Exp. I included: (*a*) the use of a different *E*; (*b*) an expansion of the recall test following the VD task so that both the right and wrong syllables of each pair were presented individually for recall; (*c*) an addition to the PA instructions of a statement that the PA task would be more difficult than the VD task, due to indications from Exp. I that *S*s may have been unaware of the difference in difficulty, and that their resulting discouragement interfered with their PA performance.

RESULTS

Since 83% of the *S*s both in Exp. I and II failed to reach the criterion of 2 successive errorless trials on the PA task within 24 trials, total number of correct anticipation responses over the 24 trials represented the most appropriate PA performance measure. This was therefore computed separately for each of the 12 PA pairs for each *S*, assuming that those few *S*s who reached criterion would have continued to respond correctly on all subsequent trials. A second measure, more appropriate as an index of initial rate of learning, was the number of errors (trials) preceding the first correct response to a given pair, designated as *before-errors*. In addition, the percentage of trials following the first correct response on which errors were made (*after-errors*) was also obtained. The means for each of these three measures are presented for the various conditions of Exp. I and II in Tables 2 and 3, respectively. Also examined was the nature of any overt errors on the PA task, which are not presented due to their infrequency and failure to show significant effects.

Although the use of different *S* populations precludes any direct comparison between the two experiments, their results will be discussed together, because of the close comparability between them in conditions and procedures. Unless otherwise stated, all comparisons reported as statistically significant refer to the .01 significance level, although all comparisons significant at the .05 level are reported as such.

D-F-N conditions (*Exp. I*).—Table 2 shows that performance for the N condition was numerically superior for all measures, with the exception of after-errors, where it was slightly surpassed by the D condition. Analyses of variance revealed significant differences between conditions for correct responses ($F = 8.34$) and after-errors ($F = 5.66$), but not for before-errors ($F = 1.75$, all *F*s based on 2 and 150 *df*). Further analysis revealed that the D and N conditions did not differ significantly from each other, but both were significantly superior to the F condition in correct responses, as well as after-errors.

D-L-U conditions (*Exp. II*).—As shown by Table 3, performance under the D condition was superior on all measures to the L and U conditions, but significantly so only for correct responses ($F = 5.23, df = 2/264$). Neither before-errors ($F = 2.69$) nor after-errors ($F = 2.35$) differed significantly between conditions. Further analysis revealed that the D condition was significantly superior in correct responses to both the L and U conditions, which did not differ from each other.

Experimental-control group (*Exp.

TABLE 2

PA PERFORMANCE MEASURES FOR CONDITIONS OF EXP. I

Measure	D Condition			F Condition			N Condition
	D(S)	D(R)	Total	F(S)	F(R)	Total	
Correct responses	10.61	8.82	9.72	7.25	8.08	7.67	9.88
Before-errors	9.29	12.25	10.77	11.31	12.01	11.66	10.28
After-errors	.279	.249	.265	.429	.326	.378	.280

II).—The performance of the control group, corresponding to the N condition of Exp. I, was numerically superior to the L and U conditions on all measures, but slightly inferior to the D condition, as shown by Table 3. However, none of these differences approached significance by analysis of variance (all *F*s <1). This lack of significance, despite the close comparability in magnitude of the differences involved with the previously described significant differences both in Exp. I and II, is a direct result of the use of different *S*s in the experimental and control groups, and the consequent necessary inclusion of between-*S*s variance in the error term appropriate for this analysis. Since the amount of total variance was quite similar within the experimental and control groups, the within-*S*s error variance from the previous D-L-U analysis involving the experimental group could be used as an estimate of error variance uncontaminated by differences between *S*s. When this was done, control group performance proved to be significantly superior to the combined L and U conditions ($F = 9.76$, $df = 1/264$), although differing at only the .05 level from the

TABLE 3

PA PERFORMANCE MEASURES FOR CONDITIONS OF EXP. II

Measure	Control Group	D Condition		
		D(S)	D(R)	Total
Correct responses	10.57	11.22	10.63	10.92
Before-errors	9.91	9.37	9.88	9.63
After-errors	.250	.233	.248	.240

L Condition

Measure	L(S-R)	L	L(S)	L(R)	Total
Correct responses	9.92	8.31	9.00	9.83	9.26
Before-errors	9.69	12.14	10.22	10.14	10.55
After-errors	.307	.300	.347	.291	.312

U Condition

Measure	U(S-R)	U	U(S)	U(R)	Total
Correct responses	9.00	9.11	9.97	9.22	9.33
Before-errors	11.72	11.44	10.17	11.22	11.14
After-errors	.267	.274	.279	.279	.274

L ($F = 5.96$) and U ($F = 6.67$) conditions separately, and not at all from the D condition ($F < 1$). While the statistical legitimacy of this procedure is suspect, it at least serves to demonstrate that the results for the N (control) condition are closely comparable between Exp. I and II, and that the apparent discrepancies with respect to statistical significance are directly attributable to the greater efficiency of the within-*S*s design of Exp. I.

Several points of agreement between Exp. I and II are noteworthy, particularly the fact that neither gave any indication of superiority for the D condition over the N (control group) condition. Moreover, all conditions wherein the syllables of the PA list were previously used in VD learning but paired differently (F, L, and U conditions) showed performance significantly below either the D or N (control) conditions. Finally, although the two experiments could not be compared directly, the difference from the D and N conditions was approximately equal for the F condition of Exp. I and for the L and U conditions of Exp. II, indicating that VD pairing with syllables from other PA pairs produces no more interference than does VD pairing with syllables not in the PA list.

VD right-wrong comparisons.—As shown in Table 2, the D(S) condition of Exp. I, wherein the stimulus of the PA pair was right on the VD task and the response wrong, was superior on all measures to the D(R) condition. However, the F(S) condition was superior to the F(R) condition only in before-errors. Analyses of variance showed that the combined D(S) and F(S) conditions were significantly superior at the .05 level to the D(R)-F(R) conditions in before-errors ($F = 5.99$, $df = 1/150$), but not for either correct responses ($F = 0.63$) or after-errors ($F = 1.98$). However, the D-F by (S)-(R) interaction proved to be significant at the .05 level for correct responses ($F = 4.74$, $df = 1/150$), which was shown by subsequent analysis to be due to the significant superiority of D(S) to D(R), while the inferiority of F(S) to F(R) was nonsignificant.

Table 3 shows that the D(S) condition was also numerically superior to D(R) for all measures of Exp. II, although not significantly so (all *F*s <1). Likewise, no consistent nor significant effect emerged from a comparison of the four possible combinations of stimulus and response, right or wrong within either the L or U conditions (these *F*s <1). With the possible exception of the L conditions, within which performance under the L(S-R) condition was best and the L condition worst, there is absolutely no evidence that PA performance is related to the number of syllables which were right on the VD task. Furthermore, in agreement with the F condition of Exp. I, PA learning was unaffected by whether the right syllable was in the stimulus or response position of the PA pair. Any such effect therefore appears to be limited to the D condition, although even here the superiority of the D(S) to the D(R) condition proved to be significant only in Exp. I.

VD recall test.—After VD learning, when *S*s of Exp. I were presented the wrong syllable from each pair and attempted to recall the right syllable previously paired with it, an average of 17% correct associations were given, and only 7 of the 36 *S*s failed to give any. In Exp. II, where this test was expanded to include the separate presentation of both the right and wrong syllable of each pair, one or more correct associations were made on 21% of the pairs for both the experimental and control groups, while only 4 of the 36 experimental *S*s and 2 of the 24 control *S*s gave none. Each group gave correct associations to 13% of the 24 individual syllables of the VD task. Presentation of the wrong syllables, while *S* attempted to recall the right ones, produced 16% correct responses, as contrasted with 11% when the right syllables were presented. This difference was highly significant (t = 5.24, df = 59), indicating that *S*s had learned more about the right than the wrong syllables during VD learning, since they were better able to recall them.

Although relatively small in number, the extent to which correct intrapair associations were given on the VD recall test appeared to warrant a comparison in performance between PA pairs involving correct associations during VD recall (associated pairs), and those which did not (nonassociated pairs). Since such comparisons require performance measures both on PA pairs which did and did not involve correct VD associations, it was necessary to eliminate for this purpose those *S*s for whom all PA pairs within a given condition were either associated or nonassociated. For Exp. I, this left 18 *S*s within the D condition and 25 *S*s for the F condition, while the D, L, and U conditions of Exp. II included 26, 19, and 13 *S*s, respectively. It was found that more correct PA responses were given to associated than to nonassociated pairs under all conditions. For the D condition, this difference was significant at the .05 level for Exp. II (t = 2.13, df = 25) and for both experiments combined (t = 2.35, df = 43), but not for Exp. I (t = 1.33, df = 17). For the combined F, L, and U conditions, the superiority of performance on associated pairs was significant (t=2.78, df=56), even though these syllables had been paired differently on the VD task, so that different associations had to be formed during PA learning. However, this difference was not significant for either the F (t = 1.93, df = 24), L (t = 1.35, df = 18), or U condition (t = 1.41, df = 12) tested separately. No meaningful statistical comparison was possible between PA pairs with one and both syllables correctly associated on the VD recall test under the F, L, and U conditions, due to the small numbers involved, although slightly more correct responses were given to such pairs with both syllables associated correctly during VD recall than to corresponding pairs with one or no syllables correctly associated.

Correlation coefficients were computed between performance on the VD recall test and total correct responses on the PA task for *S*s of the various groups, to determine whether there was any relationship between them. A significant positive relationship was found for Exp. I (r = .438), and that for the experimental group of Exp. II was nearly significant (r = .293), but little or no relationship was shown for the Exp. II control group (r = .104). Apparently formation of associations on the VD task, as measured by performance on the VD recall test, affects subsequent PA performance only when the same syllables are used on both tasks, rather than representing a general tendency for fast PA learners to form more associations during VD learning.

VD performance.—One final attempt to explicate the relationships between PA and VD learning consisted of the calculation of the number of trials and errors to the criterion of two successive errorless trials on the VD task for each *S*, which were then correlated with the various PA performance measures, as well as the VD recall test. The VD performance measures were closely comparable for Exp. I and the experimental and control groups of Exp. II. However, significant positive relationships were obtained between VD and PA performance as measured either by correct responses or before-errors only for Exp. I and the experimental group of Exp. II, with the various *r*'s ranging from .37 to .77. No significant relationships were found for the Exp. II control group, suggesting again that the use of similar material rather than a general learning ability accounts for the observed relationships between VD and PA performance. VD performance was also unrelated to scores on the VD recall test, indicating that increasing the number of presentations of the VD pairs did not affect the number of associations formed within these pairs.

DISCUSSION

The present results provide no evidence for any facilitation of PA performance by prior VD learning involving the PA material. In terms of the original rationale underlying this study, it would therefore appear that familiarization and discrimination processes are unimportant in PA learning. Such a conclusion, however, is rendered unlikely by the numerous studies demonstrating positive transfer from a variety of familiarization and discrimination procedures prior to PA learning, although such findings are by no means universal (e.g., Arnoult, 1957; Underwood & Schulz, 1960). Nevertheless, if familiarization and discrimination processes do play a significant role in PA learning, it is clear that neither was produced to any substantial degree by the present VD task.

The absence of positive transfer from VD to PA learning in this and other studies[4] (Young & Underwood, 1954), along with the present finding of no relationship between VD and PA performance within the Exp. II control group, both indicate that basically different learning processes may be involved in the VD and PA tasks. Such an interpretation receives further support from the present comparisons involving the number and position of right and wrong VD syllables in the PA pairs, especially the finding that PA performance under the D(S) condition, for which the stimulus had been right and the repsonse wrong on the VD task, was significantly superior to the reversed D(R) condition in Exp. I. That more must be learned about the response than the stimulus in PA learning has been demonstrated previously (e.g., Underwood & Schulz, 1960). Moreover, the VD recall test of Exp. II clearly showed that more had been learned about the right than the wrong syllables during VD learning. Consequently, superior performance would be expected under the D(R) condition, where the right syllables, about which more had been learned, appeared as the PA response, about which more needs to be learned. The contrasting superiority of the D(S) condition suggests that whatever is learned about syllables on the VD task must be more appropriate to PA learning of the stimulus than to PA response learning, which also points to the probable locus of any differences between the VD and PA tasks in the basic learning processes involved.

That VD learning is in fact closely compatible with learning of the PA stimulus is indicated by the fundamental importance of recognizing and differentiating between items of the list in both cases. As concerns PA response learning, at least two previous investigators have suggested that this may be basically different from VD learning. Underwood et al. (1959) concluded that their method for producing PA response learning did not involve any substantial differentiation between responses, such as would be produced by the present VD task. In a similar vein, Runquist and Freeman (1960) have hypothesized that the process of connecting the component letters of the response syllables to form a single response unit, termed "response integration" by Mandler (1954), may be largely absent from VD learning. Further evidence for the unimportance of response integration in VD learning is provided by the present finding of no relationship of VD performance to the results of the VD recall test, wherein response integration was necessary for successful performance. In contrast, response integration has previously been shown to constitute a major part of PA learning of unfamiliar material, such as the nonsense syllables employed herein (Underwood & Schulz, 1960), so that the principal difference between the VD and PA tasks may well lie in the extent to which this process is involved.

---

[4] R. K. Young, personal communication, 1960.

Differences with respect to response integration can also account for the obtained superiority of subsequent PA performance on pairs correctly associated during the VD recall test, inasmuch as successful VD recall implies that a substantial degree of response integration is present at the conclusion of VD learning, so that PA performance should be facilitated thereby. Rather than actually being produced as a result of VD learning, however, these correct VD recall associations may simply mean that the component syllables possessed greater familiarity or association value to the individual *S* prior to the experiment.

The significant negative transfer found under the F, L, and U conditions represents the most puzzling result of the present study, which defies obvious explanation. To the extent that intrapair associations had been learned on the VD task, interference of these VD associations with subsequent learning of the re-paired syllables on the PA task would offer a means of accounting for this finding, since negative transfer commonly occurs under such conditions (e.g., Porter & Duncan, 1953). However, the absence of any significant facilitating effects under the D condition, wherein the PA and VD pairs were identical, along with the relatively small number of intrapair associations shown by the VD recall test, together indicate that such associations could have been at best a minor factor in accounting for the observed negative effects on PA performance. Furthermore, performance under the combined F, L, and U conditions was actually significantly better on those PA pairs containing syllables correctly associated on the VD recall test, as compared with pairs involving no measurable learning of associations during the VD task. As suggested previously, this superior PA performance on pairs involving correctly associated VD syllables is quite likely due more to the greater prior familiarity or association value of these syllables than to the actual learning of associations during VD learning. Nevertheless, such a result is diametrically opposed to the formation of wrong associations on the VD task as an explanation for the inferiority of the F, L, and U conditions, and indicates that the primary locus of the observed interference must lie instead in the nonassociated VD syllables. Of course, the possibility remains that the VD recall test was not sufficiently sensitive to record weaker intrapair associations formed during VD learning, and the lack of any other apparent mechanism to account for the observed negative transfer effects lends credence to this suggestion. However, if the presence of negative transfer is used to infer that associations had been learned on the VD task, these associations were by definition of a partial or subthreshold variety, so that significant evidence would thereby be provided in opposition to the all-or-none theory of association formation (Estes, 1960; Rock, 1957).

Since the differences between the D condition and the combined F, L, and U conditions were as expected, but both were depressed with respect to the N (control) condition, an alternative possibility is that some general source of interference may have operated only on those PA pairs consisting of syllables used previously in the VD task. More specifically, the complex nature of the relationships of the various PA pairs to the VD task, introduced as a result of the present within-*Ss* design, may have served to confuse *S* and to interfere with his performance on these pairs as compared with the N pairs not used in the VD task. However, the virtual identity of the results of Exp. I and II, which differed primarily in this respect, argues against such an interpretation, as does a previous study wherein a within-*S*s procedure strikingly similar to the present one yielded results indistinguishable from a between-groups design where each *S* served under only one condition (Twedt & Underwood, 1959).

Although the VD and PA tasks were made to differ considerably except for the communality of the nonsense-syllable material, in order to minimize the likelihood of contamination by nonspecific sources of transfer, any negative effects produced thereby should have affected equally the N (control) condition, and therefore cannot readily account for the depression of the other conditions relative to the N condition. Therefore, while an explanation in terms of interference on those syllables common to the VD and PA tasks would be entirely consistent with the present results, there is no direct evidence to support this position, and possible mechanisms which might produce such interference remain in the realm of speculation.

In conclusion, the present study has demonstrated that prior VD learning of nonsense-syllable pairs does not necessarily facilitate subsequent PA performance on those same pairs, and that PA performance may be significantly retarded when the syllables are paired differently on the VD and PA tasks. While no satisfactory explanation is evident for either finding, and the generality of these effects beyond the present conditions remains to be determined, it is apparent that basically different learning processes may be involved in VD and PA learning, and that the further examination and specification of these differences should represent a basic and important area for future investigation in verbal learning research.

SUMMARY

In an attempt to determine the role of familiarization and discrimination processes in verbal paired-associate learning, two experiments were performed in which paired-associate (PA) learning of a nonsense-syllable list was preceded by verbal-discrimination (VD) learning, and the relationships between the syllables contained in the VD and PA pairs were systematically varied. Performance on PA pairs which were also used in the VD task (D condition) was significantly superior to conditions wherein the syllables of the PA pairs had been paired differently on the VD task, whether with other syllables not used in the PA list (F condition) or with syllables from other pairs of the PA list (L and U conditions). However, the D condition did not differ significantly from new pairs not used in the VD task with respect to PA performance, and this latter N (control) condition was significantly superior to the F, L, and U conditions. This latter negative transfer effect could not be attributed to the learning of "wrong" associations on the VD task, as measured by performance on a recall test at the conclusion of VD learning. In view of the absence of any positive effect of prior VD learning on subsequent PA performance, it was concluded that the two tasks may involve different learning processes, and some speculations were offered as to the possible nature of these processes.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

ARNOULT, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 339–350.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

HOVLAND, C. I., & KURTZ, K. H. Experimental studies in rote-learning theory: X. Pre-learning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

PORTER, L. W., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

RUNQUIST, W. N., & FREEMAN, M. Roles of associative value and syllable familiarization in verbal discrimination learning. *J. exp. Psychol.*, 1960, **59**, 396–401.

TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58**, 111–116.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. Meaningfulness and verbal learning. Chicago: Lippincott, 1960.

YOUNG, R. K., & UNDERWOOD, B. J. Transfer in verbal materials with dissimilar stimuli and response similarity varied. *J. exp. Psychol.*, 1954, **47**, 153–159.

(Received December 12, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 269–274

RELATION BETWEEN INTERSESSION INTERVAL FREQUENCY OF COMPETING RESPONSES AND RATE OF LEARNING[1]

DAVID PREMACK AND RICHARD T. PUTNEY

University of Missouri

Although several studies have shown the rate of change in performance to vary inversely with the intersession interval (ISI) (e.g., Baron, 1952; Pennington & Thompson, 1958), more recent studies have shown the frequency of certain responses to vary directly with ISI (Premack & Bahwell, 1959; Premack & Collier, 1962). Responses that vary directly with ISI are those that, instead of showing an acquisition curve over sessions, tend to have their greatest frequency on the first test session, i.e., the so-called exploratory responses. Responses of this class quite probably represent the principal competing responses of learning situations. But if the exploratory responses of one situation are the competing responses of the learning situation, then the opposite effect of ISI on learning and exploratory-response frequency, respectively, suggests that the former may be accounted for by the latter. That is, the greater the ISI, the greater the frequency of competing responses, and therefore, the slower the rate of learning.

This hypothesis was tested in the present experiment by using the light-contingent bar press as a representative competing response in an acquisition situation in which rats were required to learn to drink. By placing *S*s in the test space at different ISIs and equating water deprivation by interpolated home-cage drinking sessions, it was possible to determine the relation between bar press frequency and rate of acquisition of the drinking response.

METHOD

Subjects.—Thirty naive male albino rats of the Sprague-Dawley strain, CA 80–90 days old, were used as *S*s. They were housed individually in wood-frame cages with hardware cloth backs, fronts, and floors.

Apparatus.—Each of 10 wooden boxes, built in a rack of two tiers of 5 each, was equipped with a drinking tube and light-contingent bar. Boxes were 7 × 8½ × 8½ in., painted flat black inside and out, and had hardware cloth floors. Illumination inside the box was virtually eliminated by fitting the refuse trays close to the base of each tier of boxes.

Drinking tubes were modified 50-ml. burettes, calibrated to 0.1 ml., mounted in the center of the hinged front door through which *S* was placed in the test box. The drinking spout had an inner diameter of 5 mm., and extended ⅜ in. into the box, 2 in. above the floor. Like drinking tubes, similarly located, were used in home cage.

Each box contained an L shaped bar on the wall adjacent to the drinking tube. The bar protruded about 1 in. into the box, 2½ in. above the floor, and was activated by about 10 gm. force. Each bar press turned on a 1-in. Dialco pilot light mounted in the ceiling directly above the bar, the light remaining on as long as the bar was depressed. Bar pressing was recorded by counters housed in a remote sound-treated box.

Procedure.—Pretraining: After a week of gentling, *S*s were placed on an 11¾-hr. water deprivation schedule which was maintained throughout the experiment. Dry food was

---

[1] Part of the data reported here was taken from a Master of Arts thesis submitted by R. Putney to the Graduate School, University of Missouri. The study was begun during the senior author's tenure as a United States Public Health Service Postdoctoral Fellow and was completed with the support of Grant M-3345 from the National Institute of Mental Health.

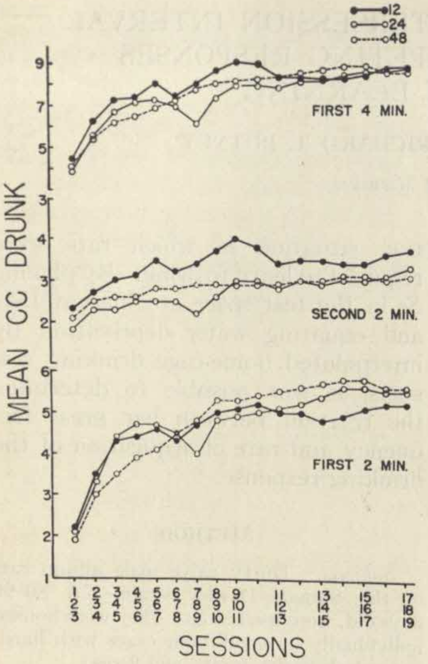


FIG. 1. Mean cc drunk averaged over running session triplets, for first 2 min., second 2 min., and first 4 min., with ISI as the parameter.

available in home cage at all times except during the drinking session, when food was removed. Drinking sessions, in both home and test cage, lasted 16 min. To insure that accommodation to the drinking schedule would not be confounded with subsequent test operations, 14 home drinking sessions were given on the deprivation schedule; these proved sufficient to stabilize home drinking. On Day 8, all *S*s were given their first drinking session in the test box. The *S*s were then equally divided into three matched groups with the matching variable being the mean proportion of the total for a session that was drunk in the first 2 min. averaged over the last four home sessions, plus the proportion of the total drunk in the first 6 min. of the first test session. The groups are referred to hereafter as 12, 24, and 48, the number denoting the number of hours that subsequently intervened between the successive test sessions which the groups received.

Testing: The experiment proper began on the morning of Day 9, with testing of the 12-hr. group. Thereafter the 12-hr. group was placed twice daily in the test boxes, at 8:30 A.M. and 8:30 P.M., and thus received no further home drinking. The 24-hr. group received water in test boxes every P.M., and in home cages every A.M., thereby maintaining the $11\frac{3}{4}$-hr. water deprivation schedule. The 48-hr. group received water in test boxes every other P.M., and had three home drinking sessions interpolated between every 2 test sessions. All groups received 20 test sessions, and after the first session each *S* was assigned to a test box in which it received all remaining sessions. Data recorded were: amount consumed in successive 2-min. intervals, both for home and test drinking; and number of bar presses by 2-min. intervals.

RESULTS

Acquisition of drinking.—Acquisition of the drinking response is shown in Fig. 1 where mean volume consumed is plotted against sessions with ISI as the parameter. Separate curves are shown for the first 2-min., second 2-min., and first 4-min. intervals, all of which have been smoothed by plotting the data quasicumulatively (see Fig. 1 for exact method). Examination of the data by successive 2-min. intervals showed that from the fourth minute on there were essentially no differences either between the groups or over sessions, and that the main consistent differences of both kinds were confined to the first 4 min. of a session. An analysis of variance of

TABLE 1

ANALYSIS OF VARIANCE FOR THE FIRST 4 MIN. OF TEST DRINKING OVER SESSIONS 1–7

Source	*df*	*MS*	*F*
Between *S*s	29		
Interval (I)	2	13.38	2.42
Error (b)	27	5.53	
Within *S*s	180		
Sessions (S)	6	161.62	118.66**
I × S	12	2.47	1.82*
Error (w)	162	1.36	
Total	209		

* $P = .05$.
** $P < .001$.

amount consumed in the first 4 min. for the first seven sessions is summarized in Table 1. The closeness of the means for the terminal sessions, 16 through 20, suggests that the groups stabilized at a common level; the significant interaction that was obtained between group and session indicates that the rate of approach to the terminal drinking level differed for the groups, being generally inversely proportional to the ISI, as shown in Fig. 1. When means for Sessions 16 through 20 are used as estimates of asymptotes, mean volumes consumed in the first 4 min. for Sessions 6 and 7 combined were: 93% of asymptote, 83%, and 77% for Groups 12, 24, and 48, respectively.

Although the groups differed in rate of approach to terminal drinking level with respect to *initial* volume consumed, there was no difference in their approach to the terminal drinking level in terms of *total* consumption. In an analysis of variance identical with the preceding one except for the use of total, rather than initial volume consumed, the Group × Session interaction failed to approach the 5% level ($F < 1.00$), and only the Session main effect was significant ($F = 20.86$, $df = 6/162$, $P < .001$). Thus the groups increased over test sessions both in initial and total volume consumed, but the rate at which they increased differed only for initial consumption.

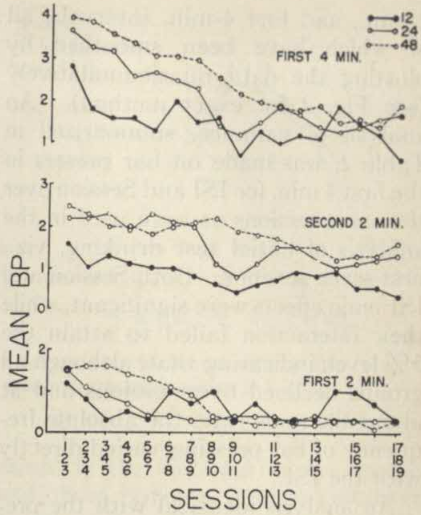


FIG. 2. Mean number of light-contingent bar presses averaged over running session triplets, for first 2 min., second 2 min., and first 4 min., with ISI as the parameter.

Interpolated home-cage drinking.—Two additional analyses established that differences in test drinking between Groups 24 and 48 were *not* due to differences in home-cage drinking that might possibly have emerged during the interpolated home-cage drinking sessions which both of these groups received during the testing proper. An analysis of variance for group and session made on the first 2 min. per session of the interpolated home-cage drinking for Groups 24 and 48 showed that neither the main effect of Group ($F = 1.96$, $df = 1/18$) nor the Group × Session interaction ($F = 1.61$, $df = 8/144$) attained the 5% level. Likewise, an identical analysis made on *total* amount drunk per session over the same period showed that neither the Group, nor the Group × Session interaction, attained the 5% level ($F < 1.00$, both cases).¶ In sum, the analyses of drinking showed that the groups differed only in initial volume consumed in test drinking, and this in terms of rate of approach to the apparently common terminal drinking level. Neither total consumption per session in test drinking, nor initial or total consumption in interpolated home-cage drinking were found to differ.

Figure 2 shows mean bar press frequency plotted against sessions with ISI as the parameter; separate curves are shown for the first 2-min., second

2-min., and first 4-min. intervals, all of which have been smoothed by plotting the data quasicumulatively (see Fig. 2 for exact method). An analysis of variance, summarized in Table 2, was made on bar presses in the first 4 min. for ISI and Session over the same sessions as were used in the analysis of initial test drinking, viz., first seven sessions. Both Session and ISI main effects were significant, while their interaction failed to attain the 5% level, indicating that, although all groups declined over sessions and at about the same rate, the absolute frequency of bar pressing varied directly with the ISI.

An analysis identical with the preceding one was made on bar presses for the session rather than for the first 4 min.; the Session main effect proved to be significant ($F=7.83$, $df=6/162$, $P < .001$), indicating that all groups declined in total bar presses over the first seven sessions, but neither the main Group effect nor the Group × Session interaction approached the 5% level ($F < 1.00$, both cases). Final analyses were made for ISI on bar presses both in first 4 min. of the session and total session summed over the last seven sessions, the large number of zero responders in these terminal sessions making a test for trend inadvisable. No differences were found either for bar presses in first 4 min. or for the session ($F < 1.00$, both cases).

TABLE 2

Analysis of Variance for the First 4 Min. of Bar Pressing over Sessions 1–7

Source	*df*	*MS*	*F*
Between *S*s	29		
Interval (I)	2	62.96	3.78*
Error (b)	27	16.66	
Within *S*s	180		
Sessions (S)	6	18.73	3.92**
I × S	12	3.41	.72
Error (w)	162	4.72	
Total	209		

* $P < .05$.
** $P < .001$.

In sum, joint consideration of the drinking and bar press analyses bears out the hypothesis of an inverse relation between frequency of the light-contingent bar press and rate of approach to the terminal drinking level. (*a*) The groups differed in rate of increase in amount drunk in the first 4 min. over the first seven sessions and, at the same time, in frequency of the bar press, there being a generally inverse relation between mean initial amount drunk and mean initial bar press frequency; also, there was an over-all increase in initial drinking concurrent with an over-all decrease in initial bar pressing. (*b*) All groups increased in total amount drunk over the first seven sessions and at the same time, decreased in total frequency of bar pressing, and there were no reliable differences between groups either in amount consumed per session or in bar press frequency per session. (*c*) All groups appeared to approach approximately the same terminal drinking level both in initial and in total consumption, and correspondingly, all groups appeared to approach about the same terminal bar press level both in initial and total frequency.

DISCUSSION

Although the present results show a rather clear inverse relation between competing-response frequency and rate of change in the referent response, it may be asked whether ISI does not have additional effects upon learning independent of competing responses. Estes (1955), for example, would derive the present results for drinking, as well as a more general body of distributional learning phenomena, from a random stimulus fluctuation assumption in statistical learning theory. Pennington and Thompson (1958), who found an

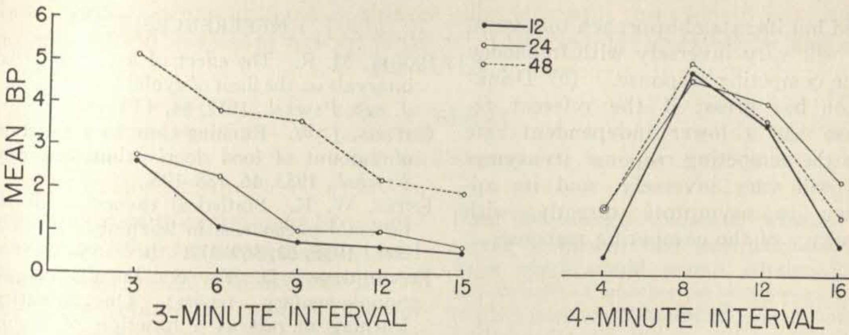


FIG. 3. Comparison of light-contingent bar pressing with and without the accompaniment of drinking. (See text for explanation.)

inverse relation between ISI and pattern discrimination learning, for ISIs identical with the present ones, ascribe the relation to a forgetting factor.

However, in none of the evidence cited for the noncompeting response interpretations of the ISI learning relation does there appear to have been any measurement of competing responses, and, of course, failure to measure competing responses does not establish that they are not present. Judged by the occasions on which competing responses have been observed in the learning situation (e.g., Cotton, 1953; Pereboom & Crawford, 1958), as well as the occasions on which their presence has been inferred to account for effects ranging from extinction to the partial reinforcement effect, it may be concluded that learning situations with, are decidedly more common than those without, competing responses. Further, though few learning situations contain the light-contingent bar press, its effect may be considered similar to that of the unmeasured competing responses that are typically present. But if it is reasonable to assume that competing responses are nearly ubiquitous, then, in order to consider whether they represent a principal, if not sole, determinant of the ISI learning relation, it need only be considered whether the positive relation between ISI and responses apt to enter the learning situation as competing responses is itself a general one. An answer to this question may be contained in the fact that in two species and three response systems, we have found no response, whose frequency is greatest on the first session, that does not vary directly with ISI (Premack & Bahwell, 1959; Premack & Collier, 1962).

The effect of the bar press on drinking is, however, only one side of the coin for there is, in addition, the effect of drinking on the bar press. Figure 3 presents a comparison of the bar press as it occurred in one case with, and in another without, the accompaniment of drinking. The right and left panels of Fig. 3 show mean light-contingent bar presses per 4-min. and 3-min. interval, respectively, averaged over the last 4 and last 6 sessions of series of 20 and 24 sessions, respectively; data in the left panel show the bar press without the accompaniment of drinking, and were taken from previous results (Premack & Collier, 1962). In addition to substantially delaying the peak bar press frequency, the presence of drinking ultimately overcame the ISI effect, so that in the terminal sessions, which these data represent, differences between Groups 12, 24, and 48 were negligible. The reciprocal effects between the bar press and drinking suggest that for at least some cases the two sides of response competition may be described by the following hypotheses. (*a*) Bar press on drinking: if the referent response has a higher independent rate (see Premack, 1959) than the competing response, its asymptote will be unaffected but its rate of approach to asymptote will vary inversely with frequency of the competing response. (*b*) Drinking on bar press: if the referent response has a lower independent rate than the competing response, its asymptote will vary inversely, and its approach to asymptote directly, with frequency of the competing response.

SUMMARY

The experiment was designed to test the hypothesis that the inverse relation between intersession interval and rate of learning may be due to a direct relation between ISI and frequency of those responses which seem most apt to enter the learning situation in the role of competing responses. In a space containing both a light-contingent bar and a drinking tube, three groups of water-deprived rats ($N = 10$) were tested repeatedly at ISIs of 12, 24, and 48 hr., respectively. Water deprivation was equated by interpolated home-cage drinking sessions, and measures were taken of both bar pressing and amount drunk. All groups approached about the same terminal drinking level, at a rate that was generally inversely proportional to their bar press frequency, which is in support of the hypothesis.

REFERENCES

BARON, M. R. The effect of long inter-trial intervals on the limit of eyelid conditioning. *J. exp. Psychol.*, 1952, **44**, 438–441.

COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

ESTES, W. K. Statistical theory of distributional phenomena in learning. *Psychol. Rev.*, 1955, **62**, 369–377.

PENNINGTON, D. F., & THOMPSON, R. Supplementary report: Discrimination learning in rats as a function of highly distributed trials. *J. exp. Psychol.*, 1958, **56**, 94–95.

PEREBOOM, A. C., & CRAWFORD, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56**, 82–85.

PREMACK, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychol. Rev.*, 1959, **66**, 219–233.

PREMACK, D., & BAHWELL, R. Operant-level lever pressing by a monkey as a function of interest interval. *J. exp. Anal. Behav.*, 1959, **2**, 127–131.

PREMACK, D., & COLLIER, G. Joint effects of stimulus deprivation and intersession interval: Analysis of nonreinforcement variables affecting response probability, *Psychol., Monogr.*, 1962, **76**(5, Whole No. 524).

(Received December 14, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 275–277

SKIN CONDUCTANCE LEVELS AND VERBAL RECALL [1]

R. N. BERRY

Indiana University

In the present experiment the question is whether there are individual differences in skin conductance levels which are related to performance on a recall test. The skin conductance level may be taken to represent "activation" in at least one sector of the organism. There is evidence that "activation," as found by a number of physiological measures, is related to intraindividual variations in performance (Freeman, 1940; Stennett, 1957). Between *S*s, however, the results could conceivably be different, since individual differences in performance measures may not be generated by differing somatic states, i.e., levels of skin conductance, and conductance, on the other hand, may be generated by factors unrelated to performance.

Malmo (1959, p. 376) has recently stated that: "It is important to stress that the measure denoted by 'moderate activation level' has meaning only in relative (not in absolute terms) terms. . . . Comparisons are invariably of the within-individual, within-task kind. . . ." This is a conservative approach, unquestionably motivated by experience and by technical considerations. However, Malmo (1958) in an earlier paper looked with favor on Schlosberg's (1954, p. 84) statement that: "Preoccupation with transient changes, the PGR, has led to general neglect of the slow drifts in absolute level of conductance, despite the fact that the absolute level is the obvious correlate of general level of tension or activation." It is not clear from Schlosberg's discussion whether he meant the statement to imply truly absolute value which would permit interindividual comparisons, or whether the intent was simply to place more emphasis on the slower changes in skin resistance.

Although the interindividual relation has not been established, neither has it been disproven, and the present experiments are a venture to test it. The type of relation found within individuals has usually been the inverted U curve showing maximum performance to occur with a moderate "activation" level. If there is an interindividual relation, one might expect it to be of this type.

METHOD

Subjects.—The 18 *S*s of Exp. I were volunteer male students from the available summer session population. All except 1 or 2 *S*s were paid on the basis of an hourly wage. The 32 *S*s of Exp. II were also male volunteers, but they were unpaid and they were taken from the author's elementary class.

Apparatus and procedure.—The following description is appropriate for both experiments. Each *S* was seated in the experimental room in a straight-backed chair with a small table in front of him. The room was separated from the recording room by a sound-reducing partition. After *S* was seated and assured that he would not receive an electric shock, the recording electrodes were attached.

The electronic circuits for the apparatus, with some modifications, are those described by Davis, Siddons, and Stout (1954). One set of electrodes provided leads to an electrical impedance plethysmograph; the data from this are preliminary, have no obvious relation to skin conductance or to performance, and are not discussed in this report. The palmar skin resistance was continuously measured by means of another ink recording amplifying system. Nonpolarizing zinc, manganese

---

[1] This research was carried out under Contract Nonr 908(03). R. C. Davis is the principal investigator under this contract. A supplementary grant was provided by the Graduate School of Indiana University. The author is indebted to R. C. Davis, K. Kveim, and F. Berry for advice and material assistance.

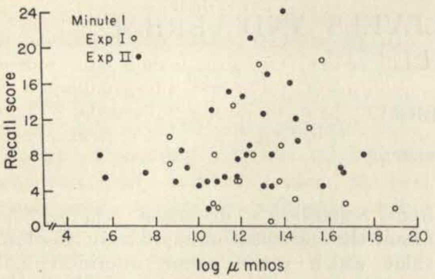


FIG. 1. The log conductance level of the first minute of the 10-min. learning session plotted against recall score.

dioxide, sodium chloride leads were placed on the left palm and upper left arm. Both skin areas were rubbed with alcohol before electrode placement, and the reference electrode on the upper arm was placed over a 7 × 1¼ in. strip of ½ in. thick foam rubber dampened with saline solution.

Following the electrode placement, *S* was given his instructions and the equipment operator made preparations to record. The *E*, in the experimental room with *S*, informed the equipment operator when the instructions were completed; the task session was then initiated.

The task.—The instructions to *S* were that he would be asked to memorize as many as he could of 30 paired associates. These were the final 30 paired items from a list by Melton and Safier, published by Hilgard (1951). Each pair had been typed on a 3 × 5 in. card and the deck of cards had been shuffled before the experiment was begun; each *S* was presented with this same order. The *S* was told, by means of sample cards that he would be required to recall the right-hand item of each of the 30 pairs when he was presented with the left-hand item. The latter was typed on another set of 3 × 5 in. cards and these had also been shuffled previous to the experiment; this shuffle determined the one order in which the cards were presented for the recall task.

The *S* was told that during the learning interval, he would have 10 sec. to study each pair and a 10-sec. intercard interval. He was further told that there would be a 6-min. interval between the learning and the recall session, and that he would be given 10 sec. to respond to each card in the recall period. All cards were presented manually on a small table in front of *S*.

Data analysis.—The midpoint values of each 1-min. interval from the beginning to the end of the task were read from the skin resistance records. These values were then converted to log conductance.

RESULTS AND DISCUSSION

All of *S*'s verbal responses were recorded during the recall period and later divided into correct and incorrect categories. The only problem in deciding about the correctness of a response were the cases where the response was not exact but was obviously some grammatical form of the correct word; these responses were scored as half correct. With this scoring procedure the range of scores was from ½ to 19 in Exp. I and 2 to 24 in Exp. II.

Following Exp. I, inspection of the data had indicated that simply dichotomizing the group into *S*s who scored better than 10 and those who scored 10 or less would be useful. All of the scores above 10 fell between the log conductance range of 1.1 and 1.6; this was defined as the optimum range. A 2 × 2 contingency table for Exp. I was set up using the above cutoff points, but Fisher's exact test showed a nonsignificant trend for the high scores to fall in the optimum conductance range.

The data of Exp. I looked sufficiently promising to justify a replication with more *S*s. The same cutoff

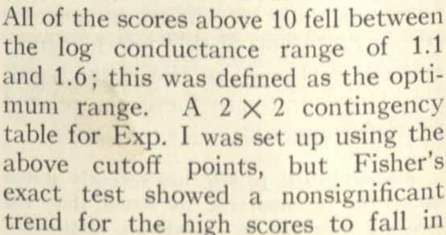


FIG. 2. The log conductance level of the first minute of the recall session (Min. 17 of the total experimental session) plotted against recall score.

points used in Exp. I were applied in a 2 × 2 contingency table for Exp. II and yielded an exact probability of observed frequencies of $P < .01$.

The log conductances for the midpoint of the first minute of the *recall* period are presented in Fig. 2. Using the same cutoffs as for Min. 1 the recall period conductance values are also predictive of recall scores; the Fisher exact test gave a $P < .05$ for the data of Exp. II. The P values for both Min. 1 and Min. 17 are one-tailed, but, since the direction of outcome was predicted from Exp. I, this is an appropriate procedure.

The important finding is that recall scores can be predicted from a measure of the organism's state in the first minute of learning. This probably represents an adjustment to the total experimental situation which is at least partially accomplished before the recording can be begun. That this adjustment persists is evident in the high correlation ($r_s = .80$ for Exp. I and .87 for Exp. II) between the conductance values of the first minute of learning and the first minute of recall.

The relation of skin resistance levels to recall scores clearly suggests the inverted U shape that has been previously found (Courts, 1942; Stennett, 1957) to relate level of activation (or arousal) to performance in complex tasks. However, it appears that optimum conductance levels do not, in many cases, insure a good performance. This latter finding must be cautiously interpreted because the resistance at the site of the reference electrode was not measured and could have varied considerably among *S*s. Since this resistance enters into the total resistance recorded, it may be that the positive finding would have been even more substantial if the reference electrode resistance had been held constant.

SUMMARY

In an effort to relate skin resistance to recall scores, two experiments, the second a replication of the first, presented male *S*s with 30 pairs of meaningful words. There were 18 *S*s in Exp. I and 32 in Exp. II. Each pair of words was presented once for 10 sec., with an interval of 10 sec. between pairs. Following this, there was a 6-min. delay before the reordered left-hand items were presented, 10 sec. at a time. The *S* was initially instructed to give as many of the right-hand items as he could remember, guessing if he wished. Skin resistance was continuously recorded throughout the experimental session and later converted to log conductance.

The data from Exp. I suggested that moderate levels of skin conductance in the first minute of the *learning* session were related to better recall. This relation was substantiated at a highly significant level ($P < .01$) in Exp. II. It was also found that moderate levels of conductance in the first minute of the *recall* period were optimal.

REFERENCES

COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347–367.

DAVIS, R. C., SIDDONS, G. F., & STOUT, G. L. Apparatus for recording autonomic states and changes. *Amer. J. Psychol.*, 1954, **67**, 343–352.

FREEMAN, G. L. The relationship between performance level and bodily activity level. *J. exp. Psychol.*, 1940, **26**, 602–608.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Ch. 15.

MALMO, R. B. Measurement of drive: An unsolved problem in psychology. In *Nebraska Symposium on Motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

STENNETT, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61.

(Received December 15, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 278–283

THE EFFECT OF STIMULUS SIMILARITY ON THE ACQUISITION AND EXTINCTION OF A CONDITIONED RESPONSE[1]

DARWIN P. HUNT

Behavioral Sciences Laboratory, Wright-Patterson Air Force Base, Ohio

The research reported here was conducted in order to test two hypotheses: (*a*) the acquisition level of a conditioned response (CR) is inversely related to the similarity between the conditioned stimulus (CS) and a nonreinforced stimulus which is presented randomly during the acquisition or training phase and (*b*) the resistance to extinction of the CR is positively related to the similarity between the reinforced and the nonreinforced stimulus.

Consider the following situation: A 50% partial reinforcement schedule is arranged in a classical conditioning situation so that a tone of 1967 cps is always followed, 500 msec. after its onset, by a shock UCS; another tone of 1000 cps, randomly presented, is never followed by the UCS. Two other groups of *S*s are similarly conditioned, with the nonreinforced stimulus being a 1967-cps or a 153-cps tone. Following acquisition, all three groups are extinguished with the 1967-cps tone.

It might be assumed that during the acquisition phase two opposing tendencies are built up, one to respond and the other not to respond, to the 1967-cps tone. Each time a CS is reinforced, the tendency to respond to that CS (and other similar ones) is increased; each time a CS is not reinforced, the tendency not to respond to that CS (and other similar ones) is also increased. Although a number of curve forms have been proposed for the generalization gradient (Shepard, 1958), it will be assumed for predictive purposes that the tendency to respond (or not to respond) is a negatively accelerated decreasing function of the psychological dissimilarity of a stimulus to the reinforced (or nonreinforced) CS (Hovland, 1937).

Within this conceptual framework, the prediction concerning the relative differences among three such groups with regard to their acquisition level seems to be clear. Using the notation of Spence (1958), and ignoring drive, i.e., assuming it to be equal for all groups:

$$E = H - \bar{I}_n$$

E represents the excitatory potential associated with the reinforced stimulus.
H is the habit strength associated with the reinforced stimulus.
$\bar{I}_n$ represents the generalized inhibitory potential associated with the reinforced stimulus.

By assuming equal terminal levels of I_n, i.e., the inhibitory tendency associated with the nonreinforced stimulus, $\bar{I}_n$ should decrease as the difference between the reinforced stimulus and nonreinforced stimulus increases. Then, if H is held constant, E would be predicted to be greater with larger differences between the reinforced stimulus and the nonreinforced stimulus.

With regard to the difference among the three groups in terms of the resistance to extinction of the CR to the 1967-cps tone, the prediction is not so straightforward. Humphreys (1939) found that resistance to extinction in a classical partial reinforcement situation was greater under a 50% reinforcement schedule than it was under a 100% reinforcement schedule. One could consider the 50% reinforcement situation as a discrimination learning task in which the nonreinforced stimulus is the same as the reinforced stimulus; while in the 100% reinforcement situation, the nonreinforced stimulus is represented by the "silence" or random stimulation occurring during the rest intervals. In this respect, then, these two situations represent the ends of a continuum of similarity of the reinforced stimulus to the nonreinforced stimulus, the 50% reinforcement situation representing the maximum similarity and the 100% reinforcement situation representing the minimum similarity between the reinforced stimulus and the nonreinforced stimulus.

---

[1] This paper is based upon a dissertation submitted to the Psychology Department of the Ohio State University. The author wishes to express his appreciation to D. D. Wickens, under whose guidance this research was conducted, for making available the GSR conditioning facilities and techniques developed under National Science Foundation Grant No. G-6203 and to W. D. Chiles for his comments and criticisms.

Secondly, the apparent degree of change involved in switching from the acquisition to the extinction phase is greater for the *S* trained under a situation having a wide difference between the reinforced and the nonreinforced stimulus. Also, wide differences between the two stimuli would increase their discriminability during acquisition, which should serve to identify the stimulus which is followed by the reinforcement. Both of these factors, i.e., a greater apparent degree of change and the ease of discrimination, would be expected to operate in the direction of lowering the resistance to extinction. These two factors suggest that resistance to extinction, as related to stimulus similarity, should directly reflect the generalization gradient.

Finally, Wickens and Snide (1955) found that the nonreinforcement of an element (a light or a tone) of a reinforced stimulus complex (a light and a tone presented simultaneously) increased resistance of the CR to extinction. The above considerations are consistent in suggesting that resistance to extinction is a decreasing function of stimulus dissimilarity, i.e., the greater the difference between the reinforced and the nonreinforced stimulus, the less the resistance to extinction.

METHOD

Subjects.—A total of 125 men and women from a beginning course in psychology served as *S*s. Of these, 25 were eliminated for the following reasons: Procedural errors (14), *S*'s skin resistance rose beyond a previously determined level during the acquisition trials (6), the intensity of the shock was too great for *S*s to continue (3), and *S* failed to respond on all reinforced trials (2). The remaining 100 *S*s provided five groups, roughly equated with regard to sex, of 20 *S*s each. The rejected *S*s were distributed approximately equally over the five groups.

Apparatus.—The apparatus consisted of equipment for the presentation of a tone and a shock to *S* and for the measurement of the magnitude of the GSR accompanying these presentations.

Any one of three tones could be selected by *E* for presentation to *S* through earphones. The three tones differed physically both in frequency and intensity, but all had a duration of 600 msec. The frequencies were 153, 1000, and 1967 cps. The intensities of the three tones were adjusted so that they were equal in loudness to 1 sone, re: mean absolute threshold, as measured at the earphones. The rates of onset and decay of the tones were controlled by an electronic switch. The rise time was reduced until a click, which would result from a too rapid onset of the tone, was eliminated. Later measurement of this rise time revealed it to be about 35 msec.

Two shock electrodes were attached to *S*'s right hand, one to the tip of the forefinger and the other to the palm. The duration of the shock was 100 msec. The time interval between the onset of the tone and that of the shock was 500 msec.

Procedure.—The 100 *S*s were assigned to one of five groups on the basis of the order of their appearance for the experimental sessions and their pretest skin resistance level. Each *S* was given 10 reinforced trials during acquisition and then 10 extinction trials. On extinction trials, as well as nonreinforced acquisition trials, the shock was omitted.

There were three experimental groups, each of which had 10 nonreinforced trials during the acquisition phase in addition to the 10 reinforced trials. These nonreinforced trials were randomly interspersed among the reinforced trials, with the following restrictions: (*a*) No more than two reinforced or nonreinforced trials could occur consecutively, and (*b*) the first and last acquisition trials (Trials 1 and 20) were reinforced and Trial 19 was not reinforced. The pattern of reinforcement-nonreinforcement was the same for all three experimental groups. The primary difference among these three groups was the similarity of the nonreinforced stimulus to the reinforced CS, the 1967-cps tone. Group E-75 was given the 153-cps tone, another Group E-25 was given the 1000-cps tone, and Group E-0 was given the 1967-cps tone as the nonreinforced stimulus. These tones were selected because they are 75, 25, and 0 jnd's, respectively, removed from the reinforced CS (Hovland, 1937).

The other two groups served as control groups and were given only the 10 reinforced trials during the acquisition phase. In Group C-M the average intertrial intervals were equal to the average intertrial intervals of the three experimental groups. In Group C-S the intertrial intervals were equal to the intervals between reinforced trials of the experimental groups.

The intertrial intervals, measured between the onsets of successive presentations of the CS, were 25, 40, 55, or 70 sec. for all groups except Group C-S. These durations were randomly assigned to the 19 acquisition intertrial intervals of the three experimental groups and to the 9 extinction intertrial intervals. The interval between the final acquisition trial and the first extinction trial was 40 sec.

Upon the appearance of each *S*, a set of standardized instructions were read to him. After placing the GSR and shock electrodes on his hands, and the earphones over his ears, the 1967-cps tone was presented alone three times at 25-sec. intervals.

Immediately following the presentations of the tone, the voltage of the shock was adjusted according to the skin resistance level such that the amperage delivered to all *S*s was generally equated. The shock was then administered three times at 15-sec. intervals. In an attempt to compensate for the adaptation of *S* to the shock, the amperage was increased from approximately 1.5 to 2.5 ma. in 0.1-ma. increments over successive reinforced trials.

The base level skin resistance was determined to the nearest 200 ohms 70 sec. after the third administration of the pretest shock. According to his base level, each *S* was assigned to one of four classes: 0–28,000 ohms; 28,200–39,000 ohms; 39,200–50,800 ohms; and above 51,000 ohms. In the final design, each of the three experimental groups and each of the two control groups contained 5 *S*s in each class. Following this assignment, the acquisition and extinction trials were administered.

Response measures.—Two measures of response were employed. *Log conductance* (LC) values were computed as given by:

$$\log\left[\left(\frac{1}{R_a}-\frac{1}{R_b}\right)10^7+1\right]$$

R_b is the skin resistance (in ohms) immediately before the presentation of the CS.

R_a is the minimum skin resistance reached after the presentation of the CS and/or UCS.

Conditioning Index (CI) scores were obtained by the equation:

$$\frac{\text{LC}}{\dfrac{\sum\limits_{t=4}^{10}\log\left[\left(\frac{1}{R_a}-\frac{1}{R_b}\right)10^7+1\right]}{7}}$$

where the numerator is the LC value for a particular trial and the denominator is the mean log conductance for a particular *S* for the last seven reinforced trials. Thus, a denominator was computed for each of the 100 *S*s. Analysis of the LC values on these seven reinforced trials for the three experimental groups indicated no difference among the groups.

RESULTS

The means of the CIs during extinction as a function of trials are shown in Fig. 1. The differences in CIs among Groups E-0, E-25, and E-75 on the first extinction trial were evaluated by the Kruskal-Wallis one-way analysis of variance by ranks test (Siegel, 1956, p. 184) and found to be nonsignificant ($.20 > P > .10$).

The experimental design is such that three main variables are involved in the extinction phase: groups, base levels, and trials. Any particular *S* was included in only one group and one base level but was administered all 10 extinction trials. This design permits an analysis of the data by Lindquist's (1956, p. 281) Type III analysis of variance design. The CIs during extinction were subjected to such an analysis. The statistical test that is crucial to the extinction hypothesis is the Group × Trial interaction. This interaction was significant ($P < .05$) whether all five groups or only the three experimental groups were included in the analysis, indicating that the rate of extinction was not the same for the different groups.

The fact that this interaction is significant and the observation that the relationship between the mean CIs and trials is approximately linear for all groups (Fig. 1) suggest that it would be appropriate and meaningful to fit a linear equation to the data for each group. A similar set of equations was computed using the LC values. The slope constants in these equations provide a quantitative estimate of the rate of extinction for each group.

The difference in linear trends among the three experimental groups based on the LC values was tested by computing a slope constant for each *S* in these groups and subjecting the resulting three arrays to Jonckheere's (1954) *k*-sample test. This yielded a *z* value of 2.17 which is significant ($P < .05$).

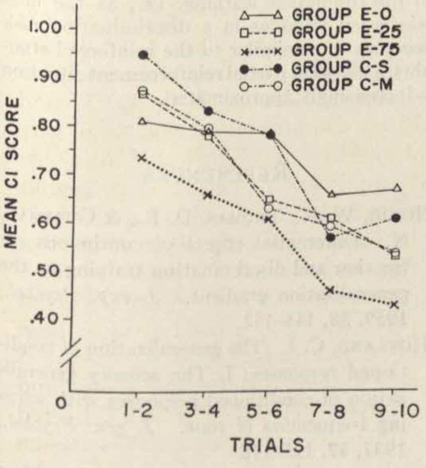


FIG. 1. The mean CI scores during extinction as a function of trials.

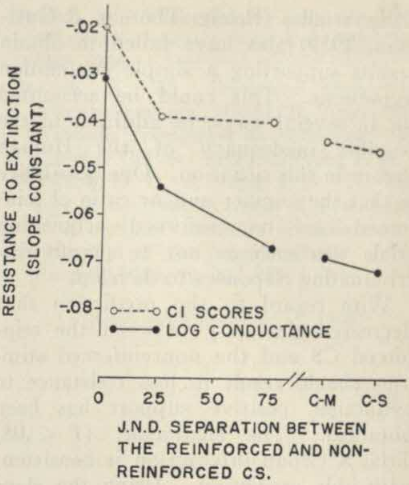


FIG. 2. Resistance to extinction as a function of the similarity between the nonreinforced and the reinforced stimulus.

These data can best be summarized by a plot showing the relationship between the slope constants and the stimulus similarity (Fig. 2).

DISCUSSION

In the initial portion of this paper two predictions were made; decreased similarity of the nonreinforced stimulus to the reinforced CS will result in (*a*) higher terminal acquisition levels of response to the previously reinforced CS and (*b*) lower resistance of this response to extinction. If the CI on the first extinction trial is an adequate index of the terminal acquisition level, and there is no apparent reason why this should not be the case, then support for the first prediction is lacking. Thus, a simple algebraic summation of these two tendencies, a positive response tendency associated with the reinforced stimulus and a negative response tendency associated with the nonreinforced stimulus, appears to be inadequate.

Other studies (Honig, Thomas, & Guttman, 1959) also have failed to obtain results supporting a simple summation hypothesis. This could be accounted for in several ways, in addition to the possible inadequacy of the Hullian theory in this situation. One possibility is that the number and/or ratio of reinforced and nonreinforced acquisition trials was such as not to permit discriminating responses to develop.

With regard to the prediction that decreased similarity between the reinforced CS and the nonreinforced stimulus should result in less resistance to extinction, positive support has been obtained. The significant ($P < .05$) Trial × Group interaction is consistent with this prediction. Using the slope constant, m, in the equation $y = mx + b$, as an index of resistance to extinction, i.e., the less negative the slope constant the greater the resistance to extinction, it can be seen in Fig. 2 that there is considerable agreement between the experimental findings and the original prediction. Using either the CI or LC values as the basis for the computation of the slope constants, results are obtained which are compatible with the original expectations. The slope constants based upon the mean LC values, however, produce a response gradient that more closely conforms to the original hypothesis than do those based upon the CIs.

The data suggest that the relative resistance to extinction of a CR to a "partially" reinforced CS can be predicted directly from knowledge of the generalization gradient involved. These results are compatible with the view that as the nonreinforced stimulus, in a discrimination learning task, becomes more similar to the reinforced stimulus, a classical partial reinforcement situation is increasingly approximated. The consequences of this are reflected in an increased resistance to extinction.

SUMMARY

This study was designed to investigate the effects of the similarity between a reinforced and a nonreinforced stimulus on the acquisition and extinction of a conditioned GSR.

Five groups of 20 *S*s each were given 10 reinforced acquisition trials consisting of a 1-sone, 1967-cps tone followed by a mild shock. Three of the groups were presented 10 nonreinforced trials randomly interspersed among the reinforced trials during acquisition. The treatment of these three experimental groups differed with respect to the similarity of the nonreinforced tone to the reinforced tone: 0, 25, or 75 jnd's separation. The other two groups were given no nonreinforced acquisition trials and differed from each other with respect to the duration of the intertrial intervals. Immediately following the acquisition phase, all *S*s were given 10 extinction trials.

The basic response measures of changes in skin resistance were transformed into log conductance values and Conditioning Index (CI) scores for purposes of analysis. The transformation into CI scores tends to compensate statistically for the general "responsiveness" of the individual *S*s.

No systematic relationship between the terminal acquisition level, as measured by the response on the first extinction trial, and stimulus similarity was found. However, a significant interaction between groups and trials during extinction was observed. The direction of this relationship was as predicted, i.e., the greater the similarity between the reinforced and nonreinforced stimulus the greater the resistance to extinction. It is suggested that this latter finding relates the concept of partial reinforcement to discrimination learning, i.e., as the nonreinforced stimulus in a discrimination task becomes more similar to the reinforced stimulus, a classical partial reinforcement situation is increasingly approximated.

REFERENCES

HONIG, W. K., THOMAS, D. R., & GUTTMAN, N. Differential effects of continuous extinction and discrimination training on the generalization gradient. *J. exp. Psychol.*, 1959, **58**, 145–152.

HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

JONCKHEERE, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

SHEPARD, W. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychol. Rev.*, 1958, **65**, 242–256.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SPENCE, K. W. Behavioral theory and selective learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958.

WICKENS, D. D., & SNIDE, J. D. The influence of non-reinforcement of a component of a complex stimulus on resistance to extinction of the complex itself. *J. exp. Psychol.*, 1955, **49**, 257–259.

(Received December 15, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 284–288

EVIDENCE FOR SHIFT EFFECTS IN THE CONSUMMATORY RESPONSE[1]

DAVID PREMACK AND W. A. HILLIX[2]

University of Missouri

The experiments reported here concern the effect of shifts in magnitude of sucrose concentration upon the mean rate of licking. The sucrose was presented directly, rather than being contingent upon a bar press, run, or choice, as in previous work (e.g., Collier & Marx, 1959; Crespi, 1942; Elliott, 1928; Young & Shuford, 1954; Zeaman, 1949); if the information sought is the effect of successive experiences with different stimuli, the tests suit the simplicity of the question to the extent that they avoid problems inherent in the instrumental response, and present the test stimuli as directly as possible. Further, the use of an instrumental response in previous work apparently diverted measurement to that response, for while previous work involved shifts in magnitudes of food or liquid, no measures appear to have been taken of *S*'s response to the food or liquid.[3]

METHOD

Subjects.—The *S*s were 36 hooded rats of both sexes, ca. 150 days old, from the University of Missouri colony. They were housed individually and maintained on ad lib. dry Purina chow and water, except for a daily 3-hr. period immediately preceding the test, when both food and water were removed from home cage.

Apparatus.—Apparatus consisted of a set of six adjoining boxes, each with a hole in the rear wall into which a drinking tube was inserted. Boxes were 7 × 7 × 9½ in., built of wood except for hardware cloth floor and rear wall, and were painted flat black on all wood sides. Drinking tubes were modified 50-ml. burettes, with apertures of about 6 mm. inner diameter. A circuit connected to each tube was so arranged that each time *S*'s tongue contacted the fluid a minute current passed through *S*'s body, activating a counter. Counter readings were taken at 1-min. intervals.

Design.—Six groups were formed by factorially combining three concentrations of sucrose (4%, 16%, 32% by weight) and drinking sessions of two durations (2 and 15 min.). The *S*s were randomly assigned to the six groups with the restriction of equal *N* and like division of the sexes.

Procedure.—Testing consisted of two 20-day periods. In Period 1, each *S* received 20 daily sessions on either 4, 16, or 32% sucrose, a session lasting either 2 or 15 min. A shift to 32% was made on Day 21, so that in Period 2 all *S*s received 20 daily sessions of 32% sucrose, a session lasting either 2 or 15 min. as before. The *S*s were run in squads of six, with each of the six conditions represented in each squad, and with concentration and session-duration counterbalanced for apparatus. Following Period 2, two additional tests were run on the 15-min. groups, and these are described separately in the results section. Sucrose solutions were prepared on alternate days from commercial granulated sugar and distilled water.

---

[1] The research reported here was begun during the tenure of both D. Premack and W. A. Hillix as United States Public Health Service Postdoctoral Research Fellows, and was completed with the support of Grant M-3345 from the National Institute of Mental Health.

[2] Present address: Navy Electronics Laboratory, San Diego, California.

[3] A few studies measure lick rate as a function of sucrose concentration, but not as a function of changes or shifts in sucrose concentration. A study by Davis and Keehn (1959) treated momentary lick rate as a function of sucrose concentration on a within-*S* basis; no differences in momentary lick rate were found. Momentary consummatory rates in the rat appear to be largely constant (e.g., Stellar & Hill, 1952); in this article only mean consummatory rates are reported, and these may be considered to reflect mainly variation in either burst length or interburst interval.

RESULTS

Total licks per session.—In Fig. 1 mean lick rate per minute averaged over successive blocks of five sessions is plotted against blocks with sucrose concentration as the parameter; the upper and lower panels are for the groups receiving 2-min. and 15-min. sessions, respectively. Total licks made in the last five preshift sessions (Days 16–20) and first (Days 21–25) and last (Days 36–40) five postshift sessions were computed for each *S*, and analyses of variance were made on these data for concentration and period. Separate analyses were made for the 2-min. and 15-min. groups and the latter is summarized in Table 1.

For the 15-min., but not for the 2-min., groups both the main effect of Period and the Period × Concentration interaction were significant. Although Fig. 1 shows the ordinal relations of the postshift (but not preshift) results for the 2-min. groups to be essentially the same as those for the 15-min. groups, no significant effects were found for the 2-min. groups, and the rest of this section concerns the 15-min. groups only.

TABLE 1

ANALYSIS OF VARIANCE OF 15 MIN. GROUPS, FOR TOTAL LICKS PER SESSION

Source	*df*	*MS*	*F*
Between *S*s	17	2,106,164	
Concentration (C)	2	777,961	.32
Error (b)	15	2,446,347	
Within *S*s	36	2,563,861	
Period (P)	2	17,247,667	18.97*
C × P	4	7,860,051	8.64*
Error (w)	30	909,085	

* $P < .001$.

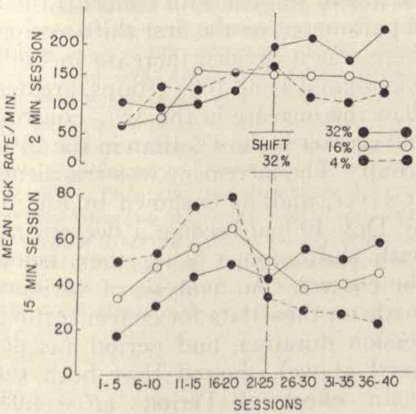


FIG. 1. Mean lick rate per minute averaged over successive blocks of five sessions with concentration (Days 1–20) and shift in concentration (Days 21–40) as parameters; upper and lower panels for groups receiving 2- and 15-min. sessions, respectively.

In the preshift period (Days 16–20), when each group received either 4, 16, or 32%, there was an inverse relation between concentration and total licks per session, and differences between all pairs of concentrations were significant at the 5% level or better. Shifting the 4 and 16% groups to 32% produced a decrement in total licks per session, the decrement being greater for the 4% than for the 16% group. In the first postshift period (Days 21–25) there were no significant differences between any of the pairs of groups, but in the last postshift period (Days 36–40) only the difference between the original 16% and 32% groups failed to attain the 5% level. Thus the decrement in total licks per session produced in the 4 and 16% groups by a change to 32% was not transient, but rather, increased from the first 5 to the last 5 postshift days.

Lick rate in first 2 min.—Comparison of the 2-min. and 15-min. session groups was made by considering licks for the first 2 min. of the session only over the same periods as in the preceding analysis. An analysis of variance was made for session-duration (D) (2- vs. 15-min. groups), concentration (C), intrasession interval

(I) (1st and 2nd min.), and period (P). The only term involving D that proved to be significant was the P × C × D × I interaction ($F = 6.6$, $df = 4/10$, $P < .001$).

The significant over-all Period × Concentration interaction ($F = 4.9$, $df = 4/30$, $P < .005$), which is plotted in Fig. 2, confirms for first 2-min. lick rates what was earlier shown for session lick rates: the 32% control increased over periods while both shift groups declined; also, the curve for the 32% control makes it clear that, despite 20 preshift sessions, asymptotic licking had not been attained when the shift occurred.

Immediate shift effects.—Although the ultimate shift effects were decremental, immediate shift effects, occurring upon *S*'s first exposure to the new sucrose values, were incremental; the increments were obscured by the preceding analyses, concerned as they were with periods effects, but are shown by an analysis of lick rate in the first minute. A stabilized estimate of first-minute lick rate for the original concentrations was obtained by averaging over the last five preshift sessions (Days 16–20); these were compared with first-minute lick rate in the first (Session 21) and last (Day 40) postshift sessions, respectively. As is shown in Fig. 3, where first-minute lick rate is plotted for the above sessions with concentration as parameter, on the first shift-session there was a decided increase in both the original 4 and 16% groups, greater than the increase in the 32% control, and greater in the 4% than in the 16% group. The increment was transient, however, and, as is shown in Fig. 3, by Day 40 had become a decrement, both shift groups being then below the control. An analysis of variance made on these data for concentration, session duration, and period (as defined above), showed that both the main effect of Period ($F = 4.03$, $df = 2/60$, $P < .025$) and the Period × Concentration interaction ($F = 3.98$, $df = 4/60$, $P < .01$) were significant, confirming that the order of the groups changed over periods.

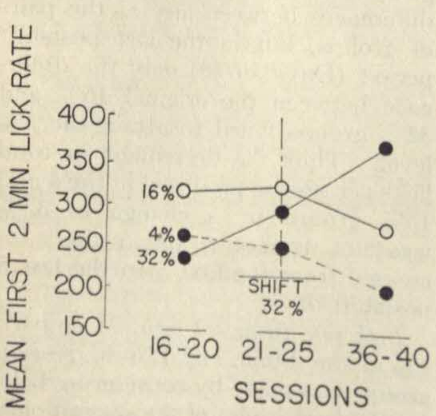


Fig. 2. Mean first 2-min. lick rate for designated blocks of five sessions with concentration (Days 16–20) and shift in concentration (Days 21–25 and 36–40) as parameters; 2- and 15-min. session groups combined.

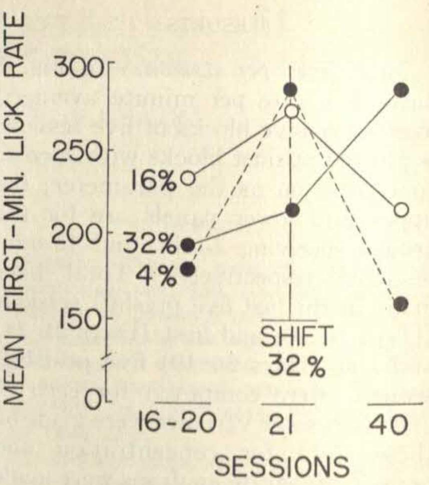


Fig. 3. Mean first-minute lick rate for designated sessions with concentration (Days 16–20) and shift in concentration (Days 21 and 40) as parameters; 2- and 15-min. session groups combined.

Permanence of shift effects.—To obtain further evidence as to the long-term stability of the shift effects, two further tests were run on the 15-min. session groups. Following Day 40, testing was discontinued for a period of 28 days. During this period, *S*s were maintained ad lib., neither food nor water being removed from home cage in the daily 3-hr. period as had been the case. Following the 28-day rest, *S*s were given 12 daily 15-min. sessions on the 32% solution as before. On the first postrest session, the groups did not differ in total licks per session, mean total licks for the original 4, 16, and 32% groups being: 546, 577, and 568, respectively. However, from the second session on the groups differed as before, mean total licks averaged over the 12 postrest sessions for the 4, 16, and 32% groups being: 381, 550, and 686, respectively. An analysis of variance on these data for total licks per session summed over the 12 sessions showed that the groups differed significantly ($F = 6.77$, $df = 2/15$, $P < .01$).

A second test was performed upon the 15-min. session groups to determine the effect of a change in food maintenance. All *S*s were changed from the previous near ad lib. maintenance to a 23-hr. food deprivation schedule. After 3 days of the feeding schedule during which testing was suspended, *S*s were given six daily 15-min. sessions on the 32% as before, the feeding schedule being maintained over the 6 test days. From the first session the groups differed as before. Mean total licks for the 4, 16, and 32% groups for the first session and averaged over the six sessions were: 729, 963, and 1,140, first session, and 893, 1,020, and 1,416, all six sessions. An analysis of variance on these data for total licks per session summed over the six sessions showed that the groups differed significantly ($F = 8.92$, $df = 2/15$, $P < .01$). Thus the follow-up tests showed that the decremental shift effect was (*a*) recoverable after 28 days of no exposure to the test stimulus, and (*b*) maintained after transfer from near ad lib. maintenance to a 23-hr. food deprivation schedule.

DISCUSSION

The results serve to establish that consummatory responding is itself subject to shift phenomena, that the mean rate at which sucrose of a given concentration is licked depends in part upon previously experienced stimulus values.

The traditional generalization concerning shift effects would appear to be that their direction, particularly, and proportional magnitude to a lesser extent, are predictable from the stimulus scale as such. That is, changes in a positive direction on the stimulus scale are thought to produce positive response changes—overshooting—and negative stimulus changes to produce negative response changes—undershooting. However, the present data do not accord with this expectation for a positive stimulus change resulted in a negative response change: groups shifted from 4 and 16% to 32% performed *below* the 32 % control.

The impression that shift effects have a directional correspondence with the stimulus scale may be the result of a generally positive correlation between stimulus values and response probabilities and, in fact, not the stimulus value, but the response probability, may be the actual basis for prediction. If so, this could be determined only in cases where the stimulus and response values were not in perfect positive correlation. Such a case is provided by the present sucrose data for, as shown in Fig. 1, there was a perfect inverse relation between sucrose concentration and mean licks for the session. In the light of this, it is of interest that the sucrose data can be described as follows: shifts from stimuli associated with higher mean response frequencies to those of lower mean response frequencies resulted in negative response changes. Though, of course, it is conjectural as to whether shift effects have a directional correspondence with a scale of response probabilities, it is nevertheless evident that traditional assumptions concerning the stimulus scale will not mediate the present sucrose results.

SUMMARY

Mean lick rate was studied as a function of changes in sucrose concentration. Rats given repeated 15-min. free drinking sessions on 4 and 16% sucrose, when shifted to 32% sucrose, showed immediate but transient incremental shift effects in first-minute lick rate, followed by long-term decremental shift

effects in both first-minute and session lick rates. The decremental effects did not abate over 20 daily sessions, and were not removed either by absenting *S*s from the test stimulus for 28 days or by a change in food maintenance.

REFERENCES

COLLIER, G., & MARX, M. H. Changes in performance as a function of shifts in the magnitude of reinforcement. *J. exp. Psychol.*, 1959, **57**, 305–309.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

DAVIS, J. D., & KEEHN, D. Magnitude of reinforcement and consummatory behavior. *Science*, 1959, **130**, 269–271.

ELLIOTT, M. H. The effect of change of reward on the maze performance of rats. *U. Calif. Publ. Psychol.*, 1928, **4**, 19–30.

STELLAR, E., & HILL, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45**, 96–102.

YOUNG, P. T., & SCHUFORD, E. H. Intensity, duration, and repetition of hedonic processes as related to acquisition of motives. *J. comp. physiol. Psychol.*, 1954, **47**, 298–305.

ZEAMAN, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received December 15, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 289–293

REACTION TIME AS A MEASURE OF INTERSENSORY FACILITATION[1]

MAURICE HERSHENSON[2]

Brooklyn College

In measuring reaction time (RT) to simultaneously presented stimuli, Todd (1912) found facilitation when sound and electric shock were paired but no effect when light was paired with either sound or shock. These results may be understood using the scheme suggested by Hilgard (1933) to account for facilitation of the eyeblink reflex. Hilgard used light and sound stimuli slightly offset in time and found that the magnitude of the eyeblink to sound was augmented when the light preceded the sound by intervals of 25 and 50 msec. Thus, he maintained, there was a summation of "reaction tendencies" when the stimuli were presented at an interval equal to the differences in the latencies of their respective responses. Since the latency for light is greater than that for sound, the light must precede the sound in order for "response tendencies" to occur simultaneously. Indeed, Hilgard suggests that the difference between the latencies for the single stimuli may be the point of maximum facilitation.

When taken together these two studies present a coherent picture. Since sound and shock yielded very similar RTs, no stimulus onset asynchrony (Δt) was necessary for simultaneity of response tendencies to result in facilitation. Similarly, facilitation should not have been expected for light and sound or light and shock when presented simultaneously since light yielded much longer RTs than either sound or shock.

The experiment to be reported was designed to study the effects of stimulus onset asynchrony on RT to paired light and sound stimuli. According to the above analysis, facilitation should be expected when the light and sound stimuli are offset in time by an amount approximately equal to the difference between the RTs to the single stimuli. The RT to this presentation should be faster than to either stimulus alone and the maximum effect should occur when Δt equals the difference between the RTs to the stimuli when presented singly. Since stimulus intensity is an important determiner of RT to single stimuli, variation of intensity might also effect facilitation. This was another variable used in the study.

METHOD

Light.—The light stimulus was produced by pulsing a glow modulator tube (Sylvania R-1131C) in the plate circuit of a 6L6 vacuum tube with rectangular pulses from a Tektronix pulse generator. The glow modulator tube was mounted in a light-tight box with a circular aperture, 1 cm. in diameter, directly in front of the tube. The light was diffused over a ground glass plate placed in front of the lamp and pressed against the inside front face of the box. The aperture, viewed binocularly from a headrest 50 cm. away, subtended 1° 10′ of arc.

---

[1] This study was based on a thesis presented to the Department of Psychology of Brooklyn College, in partial fulfillment of the requirements for the master's degree. The author is indebted to Elizabeth Fehrer and David H. Raab for their guidance and assistance throughout the course of this investigation. This research was supported, in part, by grants from the National Institute of Neurological Diseases and Blindness (B-1028) and the National Science Foundation (G-6456).

[2] Now at Yale University.

TABLE 1

MEAN RTs (MSEC.) FOR SUBJECT SR

Δt (Msec.)	Stimulus						Light & Sound Minus Δt
	Light		Light & Sound		Sound		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
0	149	11.1	98	6.3	97	6.8	98
5	144	10.6	99	5.3	94	5.5	94
15	144	8.3	111	6.3	94	7.8	96
25	144	8.5	121	6.2	94	6.4	96
35	149	7.8	131	5.9	96	7.3	96
45	149	10.1	138	5.2	97	8.0	93
55	147	8.3	141	6.6	100	5.8	86
65	153	10.5	144	4.7	104	7.0	79
75	147	8.2	142	4.2	97	8.1	67
85	145	7.8	144	8.2	99	7.1	59

Note.—Stimuli: L & S; $\bar{L} - \bar{S}$ = 49.8 msec. Each cell mean was based on 72 RTs.

Two luminances were used covering two log units. The unfiltered stimulus (L) was approximately 3000 ft-L as measured by a Macbeth illuminometer at the ground glass plate. An ND 2 Kodak filter was inserted between the glow modulator tube and the ground glass screen to produce the weaker stimulus (l). Light flashes were all 250 msec. in duration. A white cardboard ring, .75 cm. wide and 6.5 cm. in outside diameter, surrounded the stimulus aperture and served as a fixation device.

Sound.—A noise burst of 50-msec. duration was presented by closing a high speed relay with pulses from another Tektronix pulse generator. This stimulus, produced by a Grason-Stadler noise generator (Model 455B), was delivered through an audio amplifier to a set of Telephonics headphones. Two sound levels were used for both *S*s. The sensation levels corresponding to these intensities were 95 db. (S) and 75 db. (s) for Subject SR, and 85 db. (S) and 65 db. (s) for Subject MH.

Timing.—Both pulse generators were triggered from a Tektronix wave-form generator which allowed for the introduction of stimulus onset asynchrony. All critical timing functions depended on voltages which were regulated to better than 0.1%. All durations and time intervals were checked by means of a Berkeley counter (Model 7350). The RTs were measured by the counter to ±1 msec. For Subject MH, the counter was started by the light pulse for all measurements. For Subject SR, the counter was started in a similar manner except for the sound RT where the sound pulse triggered the counter. In all cases, the counter was stopped when *S* depressed a normally closed telegraph key.

Procedure.—The *S* was seated in a soundproof and lightproof room which was dimly illuminated from above and behind *S* by a 6-v. lamp enclosed in a light-tight box with a 5.5-cm. circular opening covered with several sheets of tissue paper. The luminance at the paper surface measured 22 ft-L. This lamp was 120 cm. from the ground glass plate.

Each trial was introduced by a 1000-cycle warning tone lasting 1 sec. which was delivered through a loud speaker in the chamber. This was followed by one of three silent foreperiods of 2, 2¼, or 2½ sec. selected according to a previously recorded random order. The stimulus was presented at the end of this foreperiod. The entire sequence was controlled by Hunter decade interval timers connected to present one trial every 13 sec.

A block of trials consisted of 20 reactions to a particular stimulus. Three blocks—light alone, sound alone, and the combination of light and sound with a particular preset asynchrony—made up a series. Two series were given in each session which lasted for approximately 1 hr. The sequence of the blocks was changed from series to series.

Each session began with 5 min. of dark adaptation. Four practice trials were given before each block and a 2-min. rest period was given between blocks. The *S* was instructed to respond as quickly as possible to the first stimulus perceived. Introspective reports indicated that for all asynchronies the stimuli seemed "simultaneous."

The longest and shortest RTs for each block were discarded leaving 18 RTs per block. Each asynchrony was tested over two sessions (i.e., four blocks per condition) which yielded a total of 216 RTs (72 light, 72 sound, 72 light-sound).

Subjects.—The *S*s were one undergraduate student (SR) and the author. Both were well practiced.

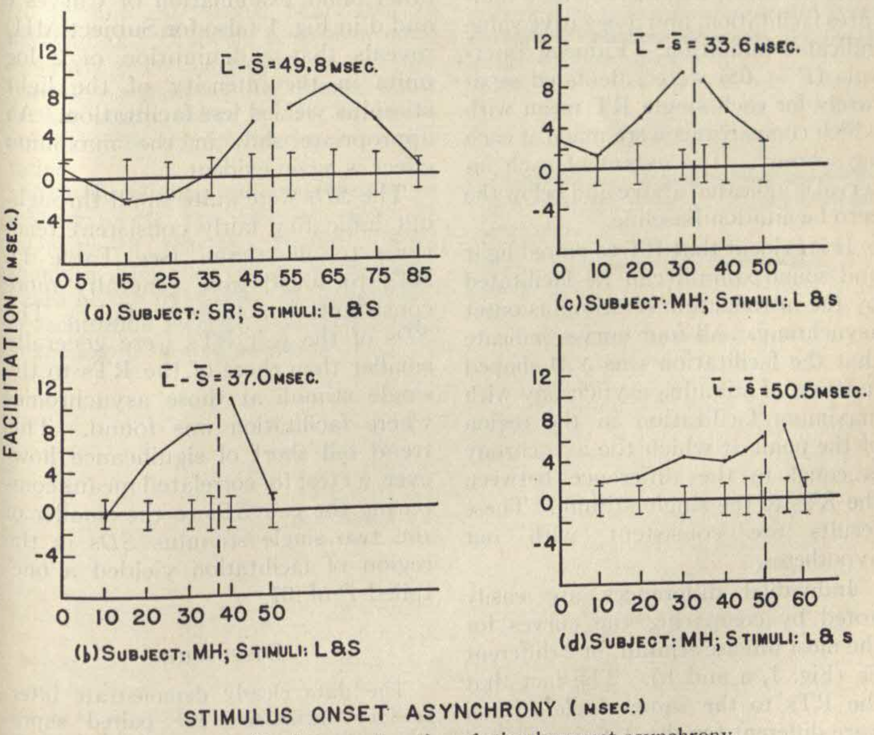
RESULTS

Table 1 contains mean RTs and *SD*s for Subject SR. The values of asynchrony indicate the amount by which the light preceded the sound for the pair RTs. In no instance did the sound precede the light.

The small differences in the mean RTs to light alone and sound alone reflect day-to-day variations. RTs to the light-sound pairs were measured from the onset of the light. If these means represented responses to the sound alone (no facilitation) we would expect the RT to increase by the amount that the light preceded the sound (Δt). The values should get progressively larger therefore, until the point where the RT to the combination equals the RT to light. The means should then remain fairly constant.

Values for RTs to the light-sound pairs minus the asynchrony indicate response latencies as if measured from the onset of the sound. Evidence for facilitation can be seen at those asynchronies where the RT to light-sound measured from the onset of the light is faster than to light alone and RT to light-sound measured from the onset of the sound is faster than to sound alone.

Figure 1 shows the curves of facili-



(a) SUBJECT: SR; STIMULI: L&S

(c) SUBJECT: MH; STIMULI: L&S

(b) SUBJECT: MH; STIMULI: L&S

(d) SUBJECT: MH; STIMULI: L&S

FIG. 1. Facilitation as a function of stimulus onset asynchrony.

tation as a function of asynchrony. These were derived from Table 1 and similar tables for Subject MH in the following way: first, overall means for all sessions were obtained for the RTs to light alone and sound alone (the best estimate of the RTs). The difference between these means is represented on the graphs as a vertical dashed line. Comparisons to the right of this line were made to light alone and those to the left of the line were made to sound alone. All comparisons were made using mean RTs which were obtained during the same sessions thereby minimizing the effects of practice and day-to-day variability.

Ordinate values were obtained simply by subtracting the RT to the pair from that to the single stimulus in accordance with the scheme outlined above. Thus a zero difference indicates no effect, a positive value indicates facilitation, and a negative value indicates inhibition. Fiducial intervals ($P = .05$) were calculated separately for each single RT mean with which comparisons were made at each asynchrony. The extent of each interval is indicated above and below the zero facilitation baseline.

It is evident that RT to paired light and sound stimuli can be facilitated by the introduction of stimulus onset asynchrony. All four curves indicate that the facilitation was a U shaped function of stimulus asynchrony with maximum facilitation in the region of the point at which the asynchrony is equal to the difference between the RTs to the single stimuli. These results are consistent with our hypotheses.

Individual differences are easily noted by comparing the curves for the most intense stimuli for 2 different *S*s (Fig. 1, a and b). The fact that the RTs to the same single stimuli were different for the 2 *S*s is reflected in the shift of the dashed line. A more important but less understandable shift was observed for Subject SR who showed maximum facilitation a full 15 msec. after the point of "physiological simultaneity." Thus for the same physical stimuli, RTs for the 2 *S*s were differentially affected by stimulus asynchrony whereas the magnitude of the facilitation was quite similar.

A comparison of Curves b and c in Fig. 1 (both curves for Subject MH) indicates that a change of 2 log units in the intensity of the sound had only a minor influence on the magnitude of the facilitation. However, the curves shifted appropriately in relation to the difference between the mean light RT and the mean sound RT, i.e., as expected, the maxima always fell near this point. On the other hand, examination of Curves c and d in Fig. 1 (also for Subject MH) reveals that a diminution of 2 log units in the intensity of the light stimulus yielded less facilitation. An appropriate shift in the maximum effect is again evident.

The *SD*s were quite small throughout indicating fairly consistent reactions to all stimuli (see Table 1). RTs to sound were generally more consistent than those to light. The *SD*s of the pair RTs were generally smaller than those of the RTs to the single stimuli at those asynchronies where facilitation was found. This trend fell short of significance however, a *t* test for correlated means comparing the pair *SD*s to the smaller of the two single stimulus *SD*s in the region of facilitation yielded a one-tailed *P* of .07.

DISCUSSION

The data clearly demonstrate intersensory facilitation for paired suprathreshold light and sound stimuli. It is further evident that this effect will be manifest only when one "gives the stimuli a chance," i.e., introduces the proper amount of stimulus onset asynchrony. Since the amount of asynchrony necessary for facilitation is determined by the difference between the means of the RTs to the stimuli when presented singly, the value will differ from *S* to *S*.

That facilitation was not due to attitudinal factors is indicated by the fact that it was a function of stimulus asynchrony and that all four curves were similar in shape. If the mere presence of one stimulus had an "alerting" or "orienting" effect on the reaction to the other stimulus, one would have expected this effect to show up at all asynchronies that were measured. Although the data cover only a small range of asynchrony, it is obvious that this was not the case.

The trend in the *SD*s may be further evidence that facilitation was not due to an alerting function of one of the two stimuli for, if this were so, it is hard to see how *SD* should be related to asynchrony. It should be kept in mind however, that this trend was not statistically significant and must await further study before it can be incorporated into our conception of facilitation.

Another finding that is difficult to understand is the different effect of intensity of the two stimuli. Why should a 2 log unit change in the luminance of the light effect the magnitude of facilitation whereas a similar change in the sound showed no effect?

There is some hint of inhibition in the data. The negative values (Fig. 1a) show that the RTs to the combination were slower than to either stimulus alone. Although this effect seems to be rather small, the values at 25 and 50 msec. asynchrony are statistically significant and indicate that this region warrants further investigation.

SUMMARY

An experiment is reported in which RTs were measured to light and sound presented singly and in combination. The chief experimental variable was stimulus onset asynchrony. The results may be summarized as follows: (*a*) Intersensory facilitation was demonstrated in that, for certain asynchronies, RTs were faster to the combination than to either stimulus alone. (*b*) Maximum facilitation occurred at or just beyond the point at which the asynchrony was equal to the difference in RT to the single stimuli. (*c*) Varying the intensity of the stimuli differentially affected the magnitude of facilitation. Reducing the light intensity resulted in less facilitation whereas reducing the sound intensity had no effect.

REFERENCES

HILGARD, E. R. Reinforcement and inhibition of eyelid reflexes. *J. gen. Psychol.*, 1933, **8**, 85–113.

TODD, J. W. Reaction to multiple stimuli. *Arch. Psychol., N. Y.*, 1912, **3**, No. 25.

(Received December 16, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 294–301

VISUAL PERCEPTION OF THE HORIZONTAL DURING PROLONGED EXPOSURE TO RADIAL ACCELERATION ON A CENTRIFUGE[1]

BRANT CLARK
San Jose State College

AND

ASHTON GRAYBIEL
United States Naval School of Aviation Medicine

Correlates of the visual perception of the horizontal and vertical have been studied in considerable detail by numerous investigators for well over 100 years (Clark, Graybiel, & MacCorquodale, 1948; Graybiel, 1952; Noble, 1949). One device which has been used in this field in recent years is the human centrifuge. In certain experiments *S* is seated on the centrifuge some distance from the center of rotation, in a fixed position with respect to the gravitational vertical, and facing the direction of rotation. If he is then subjected to a gradual increase in the velocity of rotation, he feels that he is being pressed into the seat and that the seat is being tilted away from the center of rotation. Furthermore, if *S* observes a luminous, horizontal line in the dark under these conditions, the line will appear to be displaced in the direction of the apparent tilt (Clark & Graybiel, 1951; Noble, 1949; Witkin, 1944). The amount of this apparent displacement is related directly to the deviation of the resultant force from the gravitational vertical although orientation to the resultant force does not keep pace with the physical events. This particular reorientation of the visual horizontal is a special case of the oculogravic illusion and has been observed in flight as well as on a human centrifuge (Clark & Graybiel, 1949; Clark et al., 1948).

Although this phenomenon has been known since the time of Mach (Graybiel, 1952; Noble, 1949), only two studies have been made of the oculogravic illusion during prolonged exposure to centrifugal force (Clark & Graybiel, 1951; Graybiel & Brown, 1951). In these experiments the longest period of constant rotation was 15 min., and systematic changes in the illusion related to time of exposure to change in resultant force were not found. Similar results were found for the postural vertical during flight in an aircraft by MacCorquodale (1948). In related experiments using a tilting chair, the results of experiments on exposure to tilt from the vertical are conflicting. On the one hand, Passey and Guedry (1949) found a negative aftereffect, while Pearson and Hauty (1960) reported that delay up to 60 sec. was not a factor in *S*'s judgment of the postural vertical.

It was the purpose of this investigation to study the visual perception of the horizontal on a centrifuge during periods of change in the magnitude and direction of resultant force acting on *S* for a period of 4 hr.

METHOD

Subjects.—Four experienced *S*s including the authors and two volunteers were tested. The latter were naval enlisted men assigned to the United States Naval School of Aviation

---

[1] From the United States Naval School of Aviation Medicine, United States Naval Aviation Medical Center, Pensacola, Florida. Opinions and conclusions contained in this paper are those of the authors and do not necessarily reflect the views or endorsement of the Navy Department.

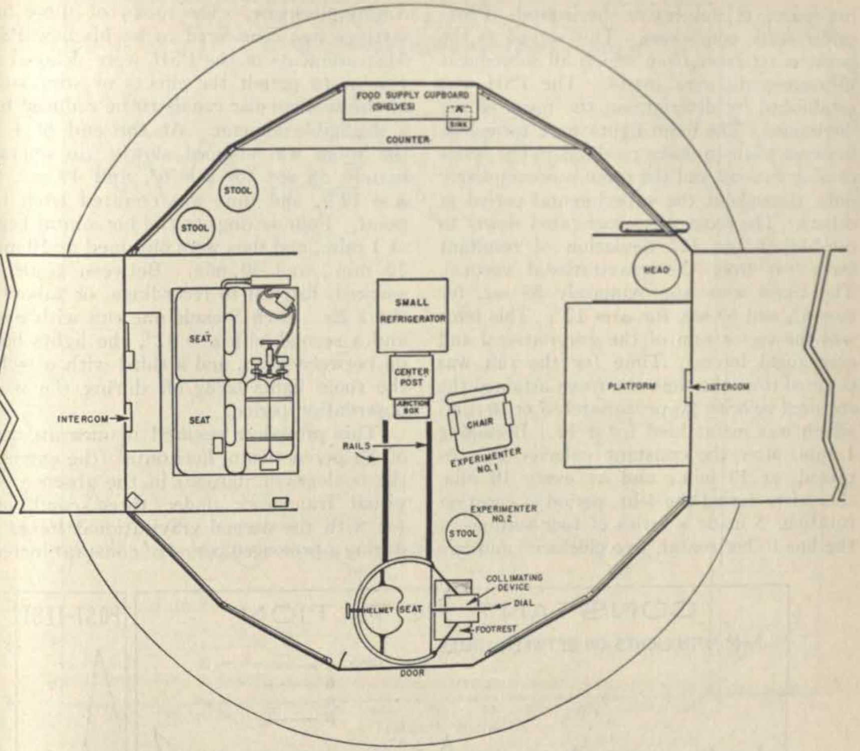


FIG. 1. Plan of the room.

Medicine with primary duty to act as *S*s in research projects. All 4 *S*s were thoroughly familiar with the problem under study and had made many other observations during previous rotations.

Apparatus.—All of the data were collected in the Pensacola slow rotation room (Graybiel, Clark, & Zarriello, 1960). This room, which is windowless and nearly circular, is 15 ft. in diameter and 7 ft. high. It is constructed around the center post of a human centrifuge, provides space for laboratory equipment, and includes living accommodations. Thus, it was possible for *E*s to be with *S*, and there was no limitation on the duration of the experimental period. The *E*s sat in positions indicated in Fig. 1, but they could move about the room at will. The *S* was seated facing in the direction of rotation in a specially constructed seat mounted firmly so that the center of *S*'s head was 6.25 ft. from the center of rotation (Fig. 1). The rotation was always counterclockwise, with deviations from the desired velocity being less than 2.5%. The seat consisted of three basic parts: (*a*) a Fiberglas head holder for each *S* molded to fit the head closely, and a chin strap to hold it rigidly in position; (*b*) a bucket seat and cloth support which could be adjusted to support each *S*'s back, the seat being adjustable vertically by a hand pump to press *S*'s head firmly into the head holder; and (*c*) a collimated luminous line of light (Clark & Graybiel, 1951) mounted on the chair directly in front of *S*. This line of light could be rotated in either direction about its center by *E* or by *S*. A luminous dial indicated the position of the luminous line and could be read by *E* to .25°. The apparatus made it possible for *S* to indicate his judgment of the horizontal by setting a line of light to horizontal in the dark during rotation, while being held rigidly in position to maintain constant throughout the testing period the relation of the head to the resultant force.

Procedure.—Each series of observations began with the centrifuge stationary. The *S* set the luminous line to horizontal in the dark 10 times, alternating clockwise and counterclockwise adjustments to establish his point of subjective horizontal (PSH) under static conditions. This served as the point of reference from which all subsequent measurements were made. The PSH was established by determining the mean of the deviations. The room lights were turned on between trials to make readings in one series of observations, and the room was completely dark throughout the experimental period in others. The room was accelerated slowly to establish 6° or 12° deviation of resultant force (ϕ) from the gravitational vertical. The times were approximately 58 sec. for $\phi = 6°$, and 69 sec. for $\phi = 12°$. This force was the vector sum of the gravitational and centrifugal forces. Time for the run was counted from the time the room attained the required velocity (approximately 7 or 10 rpm) which was maintained for 4 hr. Beginning 1 min. after the constant velocity was attained, at 10 min., and at every 10 min. thereafter during the 4-hr. period of constant rotation, *S* made a series of four settings of the line to horizontal, two clockwise and two counterclockwise. The mean of these four settings was considered to be his new PSH. Measurements of the PSH were delayed for 1 min. to permit the effects of stimulation of the semicircular canals to be reduced to a a negligible amount. At the end of 4 hr. the room was stopped slowly (in approximately 38 sec. for $\phi = 6°$, and 49 sec. for $\phi = 12°$), and time was counted from this point. Four settings to the horizontal began at 1 min., and then were obtained at 10 min., 20 min., and 30 min. Between trials, *S*s smoked, listened to recordings, or talked to the 2 *E*s. Each *S* made one run with $\phi = 6°$ and a second with $\phi = 12°$, the lights being on between trials, and a third with $\phi = 12°$, the room lights being off during the whole observation period.

This procedure resulted in measurements of *S*'s perception of horizontal (the extent of the oculogravic illusion) in the absence of a visual framework under three conditions: (*a*) with the normal gravitational force, (*b*) during a prolonged period of constant increase

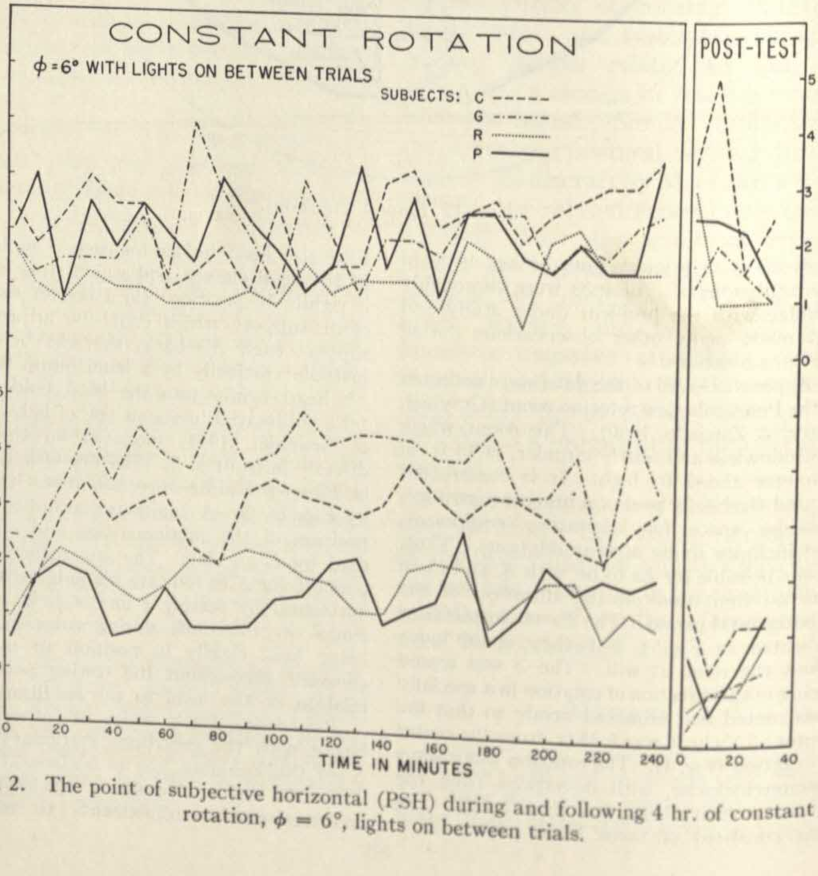


FIG. 2. The point of subjective horizontal (PSH) during and following 4 hr. of constant rotation, $\phi = 6°$, lights on between trials.

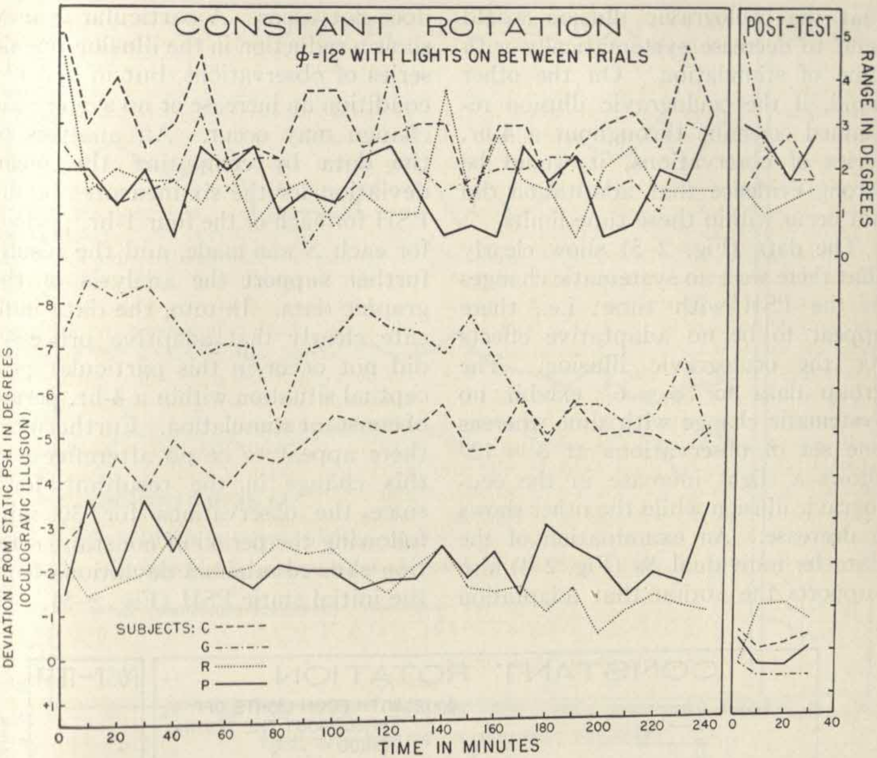


FIG. 3. The point of subjective horizontal (PSH) during and following 4 hr. of constant rotation, $\phi = 12°$, lights on between trials.

in centripetal force, and (*c*) with normal gravitational conditions immediately following cessation of rotation. Comparisons could then be made between these three conditions and the PSH (determined statically) and the two procedures used to make the measurements. In analyzing the data, evaluation of the effects of adaptation was based on time of stimulation by a constant centripetal force.

Each *S*'s point of subjective horizontal (PSH) was computed by calculating the mean of the 10 static settings of the luminous line to horizontal. This PSH served as the point from which all subsequent measurements in the series were made and is therefore the zero point of all of the graphs presented. As the rotation began, *S*s perceived the line to be displaced clockwise (considered to be plus in direction) to a new position which would remain relatively constant; i.e., *S*s observed the oculogravic illusion. In order to measure the extent of this deviation, *S* rotated the line counterclockwise (considered to be minus) to the apparent horizontal, and the mean deviation of the four settings from the static PSH was considered to be a measure of the oculogravic illusion for each point in time. Computations of median deviations also were made, but since there were few widely deviant scores, these data differed very little from the mean deviations. Therefore, plots of the mean values against time of the 4 *S*s and the experimental conditions constitute the basic data for the experiment.

RESULTS

Adaptation.—The primary purpose of this experiment was to determine whether a prolonged change in the magnitude and direction of the resultant force acting on human *S*s on a centrifuge would cause any change in the oculogravic illusion or in *S*'s PSH. If adaptive processes were involved, it would be expected that the oculogravic illusion would tend to decrease systematically with time of stimulation. On the other hand, if the oculogravic illusion remained constant throughout a 4-hr. series of observations, it would be strong evidence that adaptation did not occur within these time limits.

The data (Fig. 2–5) show clearly that there were no systematic changes in the PSH with time; i.e., there appear to be no adaptative effects for the oculogravic illusion. The group data for $\phi = 6°$ exhibit no systematic change with time whereas one set of observations at $\phi = 12°$ shows a slight increase in the oculogravic illusion while the other shows a decrease. An examination of the data for individual *S*s (Fig. 2–4) also supports the notion that adaptation does not occur. A particular *S* may show a reduction in the illusion on one series of observations, but in another condition an increase or no systematic change may occur. An analysis of the data by computing the mean deviation for the six measures of the PSH for each of the four 1-hr. periods for each *S* was made, and the results further support the analysis of the graphic data. In toto, the data indicate clearly that adaptive processes did not occur in this particular perceptual situation within a 4-hr. period of constant stimulation. Furthermore there appear to be no aftereffects of this change in the resultant force since the observations for 30 min. following the period of constant rotation showed minimal deviations from the initial static PSH (Fig. 2–5).

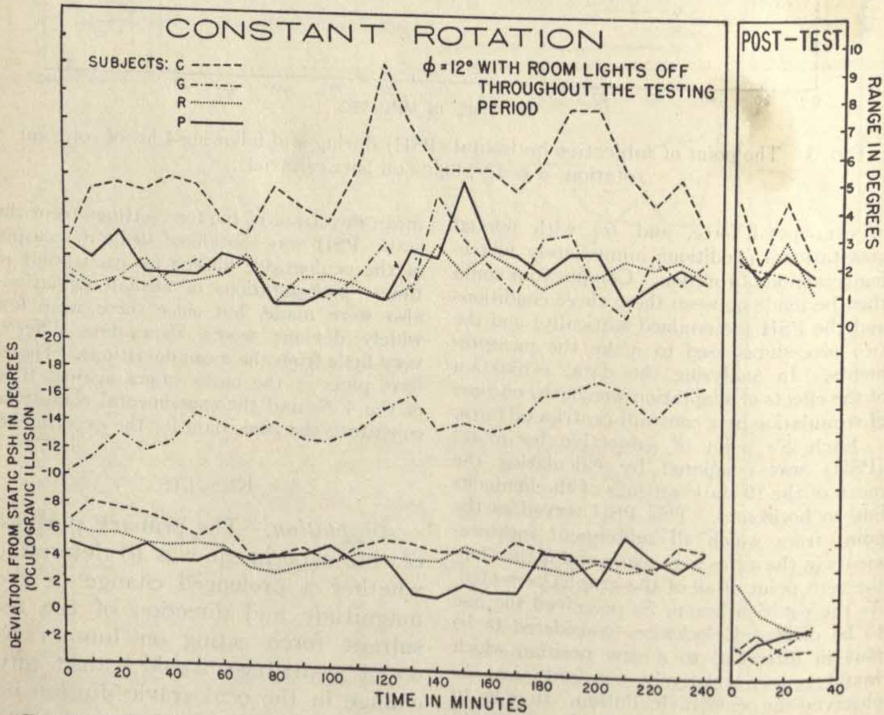


FIG. 4. The point of subjective horizontal (PSH) during and following 4 hr. of constant rotation, $\phi = 12°$, continuous darkness.

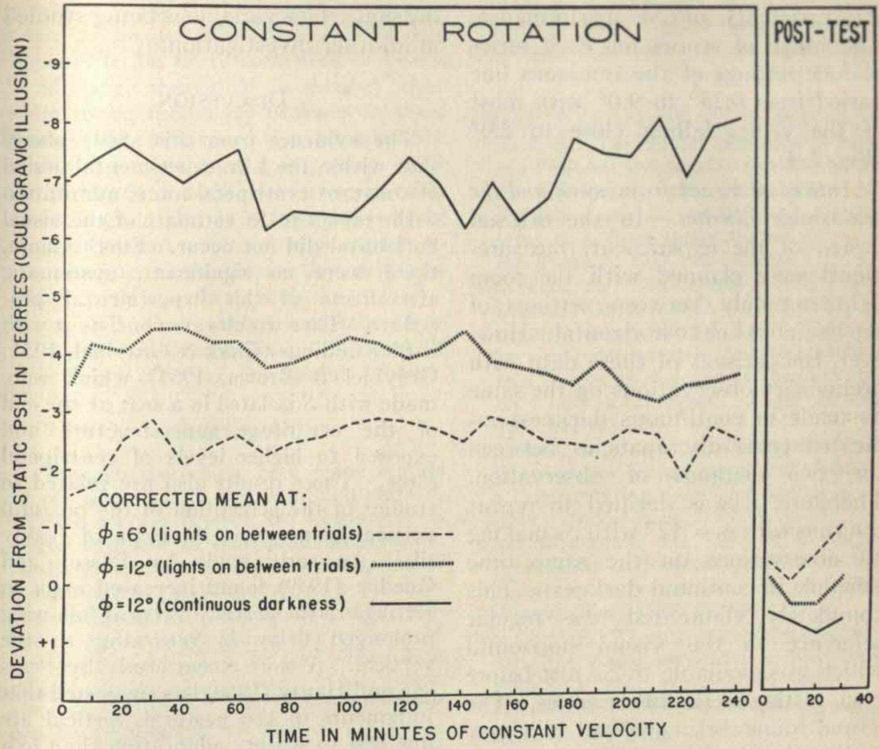


FIG. 5. The point of subjective horizontal (PSH) for 4 *S*s during the three stimulus conditions.

Relation of the oculogravic illusion to angle ϕ.—These results confirm the findings of earlier studies (Clark & Graybiel, 1951; Graybiel & Brown, 1951) which have shown that the oculogravic illusion is related to the angle ϕ, the angle between the direction of resultant force and the direction of gravity. An analysis of the group curves (Fig. 5) shows that the illusion at $\phi = 12°$ is consistently greater than at $\phi = 6°$. Similarly, the mean value for each of the 4 *S*s is greater at $\phi = 12°$ (Table 1). The results also show that the illusion measured in degrees is substantially less than the angle ϕ. The means of the 24 trials for each of the 4 *S*s beginning at 10 min. of constant rotation are given in Table 1. It will be noted that all of these values are substantially less than ϕ with the exception of 1 *S* when $\phi = 12°$ (Fig. 4). Plots of the range of scores for each series of four settings were also made to get some indication of

TABLE 1

MEAN DEVIATION IN DEGREES OF THE SETTINGS FROM THE STATIC PSH

S	Lights on between Trials		Continual Dark
	$\phi = 6°$	$\phi = 12°$	$\phi = 12°$
C	−2.96	−4.82	−5.38
G	−3.95	−6.70	−14.63
P	−1.85	−2.23	−3.96
R	−1.65	−1.69	−4.37
Mean	−2.60	−3.86	−7.09

the variability of *S*s' performance. The range of scores for each series of four settings of the luminous line varied from 0.25° to 9.0° with most of the values falling close to 2.0° (Fig. 2–4).

Antecedent visual framework and the oculogravic illusion.—In the original design of the experiment, measurements were planned with the room lights on only between settings of the luminous line to horizontal. However, comparisons of these data with preliminary observations on the same *S*s made in continuous darkness indicated gross discrepancies between the two methods of observation. Therefore, it was decided to repeat the runs with $\phi = 12°$ with *S*s making all observations on the same time schedule in continual darkness. This completely eliminated the regular reference to the visual horizontal which was available to *S*s just before each setting in the other series. The visual framework consisted of a grey wall with a horizontal aluminum bar around the room at about eye-level. The floor, ceiling, and vertical studs also were visible, but the visual field was partly obscured by the collimating device which was directly in front of *S*.

Analysis of the oculogravic illusion for the two sets of conditions for $\phi = 12°$ demonstrates that the oculogravic illusion was consistently greater when the room was continuously dark (Fig. 3–5). There is no overlapping in the two curves for the group data nor is there overlapping for *S*s G, P, and R. The fourth *S* showed overlapping of two conditions (Fig. 3 and 4), but the mean value for the final 24 measures of the PSH is greater during darkness. These data show quite clearly that a prior visual framework before a setting has a marked influence on the oculogravic illusion. This variable is being studied in another investigation.

DISCUSSION

The evidence from this study shows that within the 4-hr. experimental period of constant centripetal force, adaptation with respect to *S*s' estimate of the visual horizontal did not occur. Furthermore, there were no significant, systematic aftereffects of this experimental procedure. These results are consistent with earlier findings (Clark & Graybiel, 1951; Graybiel & Brown, 1951) which were made with *S* isolated in a seat at the end of the centrifuge superstructure and exposed to higher levels of centripetal force. These results also are related to studies of the perception of the postural vertical using short intervals of bodily tilt. An early study by Passey and Guedry (1949) found increased error in settings to the postural vertical following prolonged delay in returning to the vertical. A more recent study by Pearson and Hauty (1960) has suggested that judgments of the postural vertical are due less to sensory adaptation than to a number of other factors. The data of the present study show that the visual perception of the horizontal under the influence of centripetal force is influenced by preceding visual stimulation; i.e., the oculogravic illusion increases with a reduction in prior visual framework. This indicates that if this estimation of the visual horizontal is regarded as a special case of the oculogravic illusion, the illusion is also a function of a number of complex factors other than the input from the otolith organs.

Differences among *S*s' points of subjective horizontal for the three experimental conditions were conspicuous and consistent. The *S*s could be ranked readily not only in terms of the magnitude of the oculogravic illusion but also in terms of the individual variability and influence of prior visual stimulation. This individual consistency suggests that the 4 *S*s used different frames of reference in making the judgment. The frame of reference whether dependent on visualization of the horizontal, on otolith cues, or on proprioceptive stimulation appears to be fairly consistent for each *S*, although individual *S*s showed what seem to be temporary changes in their frames of reference by occasionally making a consistent, temporary shift in the degree of the illusion in a succession of settings. The data from this study do not present evidence with regard to the specific character of the variables which produce these individual differences, but the evidence appears clear that the oculogravic illusion is a result of a fairly complex perceptual process.

SUMMARY

Although there have been many studies on perception of the horizontal and vertical, no data are available on the effects of prolonged exposure to centripetal acceleration. In the present study of visual perception of the horizontal, 4 *S*s seated on a specially constructed seat in the Pensacola slow rotation room were studied during periods of constant rotation lasting for 4 hr. They were subjected to a constant change in magnitude and direction of centripetal force so that they perceived the oculogravic illusion.

The results show no significant, systematic changes in the oculogravic illusion during the period of constant rotation or during a 30-min. period following rotation. It was concluded that adaptation does not occur under these experimental conditions and that the oculogravic illusion is dependent upon a number of perceptual processes.

REFERENCES

CLARK, B., & GRAYBIEL, A. Apparent rotation of a fixed target associated with linear acceleration in flight. *Amer. J. Ophthal.*, 1949, **32**, 549–557.

CLARK, B., & GRAYBIEL, A. Visual perception of the horizontal following exposure to radial acceleration on a centrifuge. *J. comp. physiol. Psychol.*, 1951, **44**, 525–534.

CLARK, B., GRAYBIEL, A., & MACCORQUODALE, K. The illusory perception of movement caused by angular acceleration and by centrifugal force during flight: II. Visually perceived motion and displacement of a fixed target during turns. *J. exp. Psychol.*, 1948, **38**, 298–309.

GRAYBIEL, A. Oculogravic illusion. *Arch. Ophthal., Chicago*, 1952, **48**, 605–615.

GRAYBIEL, A., & BROWN, R. H. The delay in visual reorientation following exposure to a change in direction of resultant force on a human centrifuge. *J. gen. Psychol.*, 1951, **45**, 143–150.

GRAYBIEL, A., CLARK, B., & ZARRIELLO, J. J. Observations on human subjects living in a "slow rotation room" for periods of two days. *Arch. Neurol.*, 1960, **3**, 55–73.

MACCORQUODALE, K. Effects of angular acceleration and centrifugal force on nonvisual space orientation during flight. *J. aviat. Med.*, 1948, **19**, 146–157.

NOBLE, C. E. The perception of the vertical: III. The visual vertical as a function of centrifugal and gravitational forces. *J. exp. Psychol.*, 1949, **39**, 839–850.

PASSEY, G. E., & GUEDRY, F. E., JR. The perception of the vertical: II. Adaptation effects in four planes. *J. exp. Psychol.*, 1949, **39**, 700–707.

PEARSON, R. G., & HAUTY, G. T. Role of postural experiences in proprioceptive perception of verticality. *J. exp. Psychol.*, 1960, **59**, 425–428.

WITKIN, H. A. An investigation of problems of space orientation as related to the selection and training of pilots. *Off. Naval Res. tech. Rep.*, 1944. (Project No. N7-ONR-323).

(Received December 17, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 302–306

CONTEXT FACTORS IN PAIRED-ASSOCIATE LEARNING AND RECALL

DONALD M. SUNDLAND AND DELOS D. WICKENS

Ohio State University

The experiment reported in this paper deals with the effect of contextual cues as the discriminability of the primary stimuli is varied. Weiss and Margolius (1954) write, "It has been assumed by Guthrie, Hull, and McGeoch that responses can become associated not only with primary stimuli but also with context stimuli" (p. 318). In their experiment it was found that learning required fewer trials if there were differently colored backgrounds for the paired associates rather than uniform gray backgrounds. They also found that the recall, for groups learning under the colored backgrounds condition, was better when the color alone was shown than when the words were shown on uniform gray backgrounds. The context cues were more useful than the primary stimuli. The present experiment was designed to further clarify the role of context stimuli.

It was hypothesized that where the primary stimuli were highly discriminable, the addition of discriminable contexts would not facilitate the learning task. A corollary to this was that when the primary stimuli were less discriminable the addition to discriminable contexts would facilitate the learning task. The latter is merely a restatement of one of the results of Weiss and Margolius. It was also hypothesized that where the primary stimuli were highly discriminable, the recall to the associated discriminable contexts will be less than when the primary stimuli are poorly discriminable.

METHOD

Experiment I

Subjects.—The 114 *S*s were female students, aged 18 to 20 yr., from an elementary psychology course, who chose to participate in this experiment to meet a requirement of the course. They were tested in 10 groups ranging in number from 8 to 16 *S*s.

Apparatus.—A Patterson memory drum was fitted under an opaque projector. The image of the drum was projected onto a screen in a semidark room. The *S*s were provided with booklets in which to write their responses. It consisted of lines on which the response could be written. There was a separate page for each trial. The lists were printed on white window-shade material in black India ink. The colors were painted on using "Permanent Pigments" water colors with a liquid detergent in the water, since otherwise the paint did not spread on the window-shade material. One set of primary stimuli were six-letter words selected from the first 1,000 most frequent words as listed in Thorndike (1944). The other set of primary stimuli were two-syllable, hyphenated, nonsense syllables, each chosen from the Glaze (1928) list of 7% association value. The responses were three-letter words selected from the first 100 most frequent words as listed in Thorndike (1944).

An attempt was made to avoid logical connections between the stimulus and the response and between the colored context and the response. The nine paired associates and the three rest interval pairs (TURN-NOW, NOW-REST, START-NOW) were printed on a circular tape with five replications, the order of the paired associates being randomized. Table 1 presents the responses, the stimuli, and the colored contexts.

Procedure.—The *S*s were told that they were to attempt to anticipate the appropriate response by writing it in their booklets and that a new page would be used for each trial. After Trials 5, 10, and 15 the stimuli only were shown. These are called the test trials. Following each test trial a 30-sec. rest interval was introduced, during this time *E* circulated through the room and looked at the booklets to check whether *S*s had understood the instructions. An *S* was discarded if he responded to the rest interval pairs, or if *S* had written down all of the responses from the second learning trial on since it is unlikely that *S* could have learned the list in one trial and this would indicate he had misunderstood the instructions. The stimulus was shown alone for 2 sec. and the S-R pair was then shown for 2 sec. Two groups were tested under each of five learning conditions: color only (C); words only (W); nonsense syllables only (NS); words on colored backgrounds (C+W); one nonsense syllables on colored backgrounds (C+NS). After the regular test following Trial 15, the C+W and C + NS groups had two additional test trials. One group under each of these two learning conditions was tested first with the C list and then with the W (NS) list. The other group was tested in the reverse order.[1]

TABLE 1

STIMULUS AND RESPONSE MATERIALS

Response	Stimulus Items		
	Words	Nonsense Syllables	Colors
ARM	SPIRIT	ZIH-MIF	GREEN
ACT	FLOWER	FUH-YIG	UMBER
COW	SUDDEN	ZUB-YEK	ROSE
BAG	MANNER	QAS-ZEG	GREEN
OAK	DINNER	DAQ-GUK	VIOLET
WAY	BRANCH	BOF-NUX	YELLOW
CUP	DEGREE	CEJ-KEB	RED
DOG	MINUTE	MIB-VUF	ORANGE
OIL	LENGTH	GEC-ZOS	BLUE

Experiment II

This experiment was undertaken to obtain additional data which might facilitate an understanding of the factors which were operating to produce the results obtained in the first experiment. It was conducted in precisely the same manner and with the same materials and type of *S*s as in Exp. I, but included only two major groups of 28 *S*s each, color plus words, and color plus nonsense materials. In the retention test half the *S*s were tested to the color first and half to the symbol first. Following the last retention condition they were asked to describe the nature of the stimulus they had employed during the learning. Four possible choices were given to *S*s: symbols only; colors only; symbols and colors together; symbols sometimes and colors sometimes. The *S*s who chose any stimulus using colors were requested to write down the names of the colors they used.

TABLE 2

MEAN CORRECT RESPONSES ON TEST TRIALS 5, 10, AND 15: EXP. I

Cond.	*N*	Test Trial					
		5		10		15	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
C	30	3.50	2.23	5.21	2.21	6.87	1.86
W	18	4.67	3.06	7.05	1.81	7.72	1.72
NS	24	2.96	2.19	6.13	2.01	6.88	1.88
C + W	18	4.16	2.52	6.94	1.87	8.33	0.95
C + NS	24	2.83	2.19	5.88	2.01	6.96	1.88

RESULTS

Experiment I

Tables 2 and 3 summarize the results of this experiment. The index of learning used was the number of correct responses on the test trials. The degree of recall was indicated by the number recalled correctly divided by the number of correct responses on the last regular anticipation trial, this ratio being expressed as a percentage.

TABLE 3

PERCENTAGE RECALL TO ELEMENT: EXP. I

Recall Order	*N*	Symbol		Color	
		Mean	*SD*	Mean	*SD*
C+W→W→C	12	100	11.6	6.7	10.8
C+W→C→W	12	95.5	8.2	14.9	13.8
C+NS→NS→C	12	77.8	31.6	37.8	34.5
C+NS→C→NS	12	67.8	19.5	56.7	31.9

---

[1] We are grateful to Gloria Barts for her help in preparing the experimental materials and in collecting the experimental data. The research was supported in part by NSF Grant G6203 to the junior author.

TABLE 4

Analyses of Variance of Recall Scores on Test Trials 5, 10, and 15: Exp. I

Source	*df*	Trial 5 *F*	Trial 10 *F*	Trial 15 *F*
Between *S*s	3			
Context (C)	1	.24	.16	.57
Stimuli (S)	1	6.79*	4.72*	2.62
C × S	1	.10	.00	4.90*
Error	80			
Total	83			

* $P < .05$.

That *S*s were able to learn the list under Cond. C proves that the colored backgrounds were discriminable and sufficient for the learning task. Ten of 30 *S*s in Cond. C had all correct on the last test trial.

Since the groups varied in number, *S*s were discarded randomly until the cells of a Context × Stimuli design contained proportionate numbers. Table 4 shows the results of separate analyses of the three learning test trials. At no trial did the addition of colored contexts effect the learning performance. For Trials 5 and 10 Groups W and C + W attained higher scores than Groups NS and C + NS. At Trial 15 there was a significant interaction between the context and stimulus variables. However, this was in the unpredicted direction, Group C + W behaving as if the colored contexts added to their performance.

TABLE 5

Percentage Recall to an Element: Exp. II

Recall Order	*N*	Symbol		Color	
		Mean	*SD*	Mean	*SD*
C+W→W→C	14	101.4	12.3	10.3	16.0
C+W→C→W	14	81.6	35.7	8.7	15.7
C+NS→NS→C	14	73.5	33.3	39.3	34.9
C+NS→C→NS	14	82.8	37.4	36.5	35.9

The second hypothesis implied that the use of the context cues alone would be more disturbing to the word group than to the nonsense group. The data was analyzed by means of a $2 \times 2 \times 2$ analysis of variance. The columns and rows of the design corresponded, respectively, to the learning stimulus conditions (C + W vs. C + NS) and the test conditions (context alone vs. stimulus alone). The third dimension of the design represented the order of testing. Of the three main effects, only the test conditions (color or symbol) affected the degree of retention, *S*s showing more retention to the symbol ($F = 115.01$, $df = 1/88$, $P < .01$). Of the four interactions, only S × T was significant ($F = 34.25$, $df = 1/88$, $P < .01$). As hypothesized, Group C + NS was able to give more correct responses when presented with the colored contexts alone than Group C + W. The recall percentage for the word groups to the context cue alone was only 10.8% as opposed to 47.8% for the nonsense groups ($P < .001$). To the symbol alone, however, the word groups excelled

TABLE 6

Percentage Recall to an Element According to the Stimulus Used: Exp. II

Route	*N*	Symbol		Color	
		Mean	*SD*	Mean	*SD*
C + W	28				
W only	26	90.6	29.7	6.8	12.6
Sometimes	2	100	.00	44.5	11.5
C + NS	28				
NS only	10	100.4	12.1	1.7	5.1
NS and C both	6	89.5	16.3	54.3	12.9
Sometimes	8	77.3	27.3	49.9	11.1
C only	4	5.5	5.5	79.0	18.5

the nonsense groups, with 97.8% and 72.8%, respectively ($P = .001$).

Experiment II

The results for the recall conditions are presented in Table 5. An analyses of variance gave essentially the same results as were obtained in Exp. I in that: (*a*) *S*s showed a higher degree of retention to the symbols than to the colors ($F = 111.86$, $df = 1/104$, $P < .01$); (*b*) Group C + NS showed a higher degree of retention to colors than did C + W; (*c*) Group C + W showed a higher degree of retention to the symbols than did C + NS (for Stimulus ×Test Element interaction, $F = 13.18$, $df = 1/104$, $P < .01$).

To probe more deeply into the factors which resulted in these retention differences, the groups were divided on the basis of *S*s' stated stimulus preferences during learning. The data are shown in Table 6. It would seem that for Group C + W the vast majority of the *S*s reacted to the symbols. However, in Group C + NS the situation was quite different and the majority of the *S*s either supplemented the symbols with the colors or they used the colors alone.

For the 18 *S*s of Group C + NS who used color in some way, a rank order correlation was run between the number of colors named and the percentage recall to color. The magnitude of the rho was .84 ($P < .001$).

DISCUSSION

The results seem to indicate that the effect of context cues is a relatively complex one. There is no evidence in the data that the context cue facilitates learning, regardless as to whether the symbolic cue is a meaningful word or a nonsense syllable. This is true despite the fact that the context cue alone was as adequate a stimulus for learning as the nonsense syllables—that is, nonsense syllable alone, color alone, and nonsense syllable plus color are essentially the same. The results for words with and without color follow a similar pattern. When, however, the symbolic cue is removed for those who have learned with both present, the word group does extremely poorly; the nonsense group does not deteriorate to so marked an extent. Removing the context cue, however, has little, if any, effect upon the word group although it seems to depress the nonsense group slightly.

The results for the word group do not seem to be difficult to understand. The context cue seems to be essentially irrelevant to their behavior; it does not facilitate their learning, and it is essentially useless to them in the absence of the word. Apparently the highly discriminable verbal symbol dominates the stimulus situation during learning and suppresses the potential effectiveness of the context cue. It is as if these *S*s failed to respond to the context cue as a potentially helpful stimulus during the learning. The verbal reports obtained in Exp. II do indeed indicate that the vast majority of these *S*s reported using only the verbal symbols as stimuli.

When the symbolic stimulus is a nonsense syllable, however, the context cue acquires sufficient habit strength so that it alone can serve as a fairly effective stimulus in the retention-transfer situation. It might be expected that the presence of this additional and potentially effective dimension of differences among the stimuli would have facilitated performance during the learning of Group NS + C, but there is no evidence that such is the case. The verbal reports suggest an explanation of this result. Only 6 *S*s stated that they used both dimensions of the stimulus, the others reporting that they used one or the other alone, or sometimes one and sometimes the other. In short, most of them behave as if only a single dimension were available at each S-R pair. The learning data offer some slight but nonsignificant evidence that the readiness to use both or either dimension may have had a facilitating effect, for when the color or nonsense syllable alone *S*s were combined and contrasted with the both or sometimes *S*s, the mean for the dual *S*s on Trial 15 was 8.07 as compared with 7.57 for the single *S*s. Such a result, if real, could arise from differences in ability levels of *S*s who are alert enough to attempt to make use of both dimensions as contrasted with those who chose only one, as well as to a difference in the efficiency of the two methods. A different type of experimental design from that employed in the present experiment will be required to answer this question as well as to discover the conditions which are more favorable to the utilization of both dimensions of the stimulus.

Weiss and Margolius (1954) had found that context cues alone were more effective response guides than were the symbolic cues when they were slightly modified. In the present experiment the context cues proved to be almost completely ineffective as sole cues in the situation where they had been combined with words, but fairly effective when employed with nonsense syllables. In both their experiment and in the present one the context cues alone could be used as stimuli for the learning.

The present experiment does not in any essential way contradict the Weiss and Margolius (1954) study, but rather can be combined with it to extend our knowledge concerning the operation of context cues. Briefly, the experiments suggest that effectiveness of context cues will be a function of an interaction between the discriminability of the major cues among themselves and the discriminability of the context cues among themselves. This experiment has only shown that context cues will have greater or lesser cue value as a function of the nature of the symbolic stimulus, but it is plausible to assume that the same relationship would hold if context discriminability were varied and symbolic cues held constant.

SUMMARY

An experiment was conducted investigating the learning of paired associates, with and without differently colored backgrounds, and with "high frequency" words and low association value nonsense syllables as stimuli. The various lists were learned by 114 *S*s in groups, using a memory drum attached to an opaque projector.

It was found that the addition of differently colored backgrounds did not increase the ease of learning either where the stimuli were high frequency words or where they were low association value nonsense syllables. The associations between the colored backgrounds and the responses were better reported when the nonsense syllables were the original primary stimuli than when the words were. The experiment suggests that effectiveness of a context cue will be a function of an interaction between the discriminability among the individual primary cues and the discriminability among the context cues.

A subsequent experiment was conducted in which *S*s were asked to describe the stimuli they used in learning the list. Their verbal reports suggested that the results could be attributed, at least in part, to a tendency to select only one of the two available cues.

REFERENCES

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

THORNDIKE, E. L. *The teachers workbook of 30,000 words.* New York: Teachers College, Columbia University, 1944.

WEISS, W., & MARGOLIUS, G. The effect on context stimuli on learning and retention. *J. exp. Psychol.*, 1954, **48**, 318–322.

(Received December 23, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 307–313

TESTS OF THREE HYPOTHESES ABOUT THE EFFECTIVE STIMULUS IN SERIAL LEARNING[1]

ROBERT K. YOUNG

University of Texas

The items of a serial task are presented in a constant order; because of this the effective stimulus for a specific response may possibly be one of several alternatives: It may be: (*a*) the item immediately preceding the response to be learned; (*b*) a combination of several items preceding the response to be learned; (*c*) the position which the item holds in the list; (*d*) some combination of these; or (*e*) something else entirely.

The present paper reports a series of three experiments designed to help identify the effective stimulus in serial verbal learning. The first alternative (*a*), called the specificity hypothesis (Young, 1961), is tested in Exp. I by measuring transfer from a serial list to a subsequent paired-associate (PA) list composed of pairs taken from adjacent items in the serial list. If associations are formed between adjacent items in serial learning, positive transfer should obtain from the serial to the PA list and the PA transfer pairs should be learned faster than control pairs which have not been previously learned in serial order.

The second alternative (*b*), called the compound-stimulus hypothesis, is tested in Exp. II by measuring transfer from a serial list to a PA list composed of items taken from the serial list. Positive transfer from the serial to the PA task would be taken as evidence favorable to a compound-stimulus interpretation of the effective stimulus in serial learning.

The third alternative (*c*), called the serial-position hypothesis, is tested in Exp. III by observing transfer from one serial list to a second composed of the same items as the first. The serial position hypothesis assumes that items holding the same serial position in both lists (S items) would show positive transfer from the first to the second list while items holding different positions in both lists (D items) would show negative transfer from the first to the second list. The reason for this is that S items constitute an A-B, A-B paradigm of positive transfer in which the stimulus (A) would be the constant serial position, and the response (B) would be the same item from list to list. The D items would constitute an A-B, A-Br paradigm of negative transfer (Porter & Duncan, 1953) in which the response to be learned to a constant position stimulus would be rearranged from the first list to the second.

GENERAL METHOD

In the three experiments all *S*s were selected from introductory psychology classes at the University of Texas. All lists were constructed of dissimilar adjectives taken from Melton (Hilgard, 1951), and all were learned by the anticipation method. The lists were displayed by a filmstrip projector. The PA lists were learned at a standard 2:2-sec. rate with a 4-sec. intertrial interval while the serial lists were learned at a standard 2-sec. rate with a 6-sec. intertrial interval. When serial and PA lists were learned in an experiment, a 5-min. interval was employed

---

[1] The research reported in the present paper was supported in part by the National Science Foundation (G-4850) and in part by a grant from the Graduate School of the University of Texas.

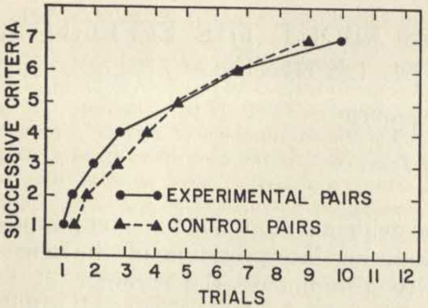


FIG. 1. Trials to successive criteria curves for the experimental and control sublists during PA learning in Exp. I.

between lists to allow time to change film strips and to read PA instructions; and when two serial lists were learned, a 1-min. interval between lists was needed to change film strips. All lists (with the exception of Exp. Ia) were learned to one errorless recitation and all PA lists had four orders of presentation to reduce the possibility of serial learning.

EXPERIMENT I

Method.—Each *S* learned a serial list and then a PA list. One-half the items in the PA list were experimental pairs obtained from adjacent items in the previously learned serial list, and the other half of the PA items were control pairs. The ordering of the experimental pairs in the PA list was exactly the same as in the serial list. That is, for these pairs the same stimulus elicited the same response in both serial and PA learning.

To reduce the possibility of differences due to intrinsic differences in difficulty in the pairs themselves, the items which served as experimental pairs for half the *S*s served as control pairs for the other half. Subgroups of 18 *S*s learned one of two serial lists and then the common PA list containing items from both serial lists (Total $N = 36$). Two 14-item serial lists (plus a cue symbol) were constructed and one 14-pair PA list was derived from the two lists.

Results.—The PA list may be thought of as being composed of two seven-pair sublists. The mean trials to learn the experimental and control sublists to one errorless recitation were 10.06 and 9.06, respectively ($F = .97$, $df = 1/34$). The mean correct responses per *S* during learning for the experimental and control sublists were 62.58 and 63.22, respectively ($F = .17$, $df = 1/34$).

As in a previous experiment measuring transfer from a serial to a PA list (Young, 1959), positive transfer occurred early in learning and then disappeared later in learning. The mean number of trials to attain successive criteria are plotted in Fig. 1. The *F*s obtained between the experimental and control sublists for criteria of one, two, three, and four correct responses ($df = 1/34$ in each case) were 4.86 ($P = .05$), 5.25 ($P = .05$), 10.50 ($P = .01$), and 7.61 ($P = .01$), respectively. Beyond a criterion of four correct responses, the two sublists do not differ with respect to mean trials to successive criteria.

The associations of the experimental pairs were differentially reinforced during serial learning. For example, Item 2 in the serial list was correct .84 times per trial; in contrast, Item 10 was given .39 times

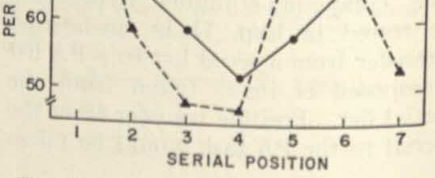


FIG. 2. Mean percentage correct in PA learning for the experimental and control sublists in Exp. I with the responses being ordered along the baseline as they were in the serial lists.

per trial. This difference reflects, of course, the bowed serial-position curve commonly found in serial learning. Since adjacent pairs were taken from the serial list and used in the PA learning, serial position effects should transfer to the PA learning if associations are formed between adjacent items. Where the pairs in the two PA sublists were arranged according to position in their respective serial lists, the mean percentage frequencies with which these experimental and control pairs were correct during PA learning are shown in Fig. 2. Of the seven comparisons made, one significant difference was obtained. As can be seen in Fig. 2, the last pair in the experimental sublist was given correctly more often than its control ($F = 28.73$, $df = 1/34$, $P < .001$).

The PA error data were analyzed in two ways. The stimuli in the PA list can be classified either as experimental or as control, and the data were analyzed according to this classification. The mean numbers of errors per trial elicited by experimental and control stimuli were .52 and .39, respectively. This difference is significant ($F = 5.87$, $df = 1/34$, $P < .05$). Secondly, the mean numbers of incorrect experimental items per trial elicited by an experimental stimulus and the mean numbers of control items per trial elicited by a control stimulus were .44 and .32, respectively ($F = 7.05$, $df = 1/34$, $P < .05$). It is noteworthy that the remainder of the errors per trial, .08 and .07, is small and constant across conditions.

EXPERIMENT Ia

The associations formed during serial learning may not have been of sufficient strength to transfer to the subsequent PA task. Experiment Ia was conducted to control for this possibility. The design of this experiment is exactly the same as Exp. I with one exception: the serial learning criterion was increased to one errorless recitation plus an additional 10 trials. A total of 22 *S*s were employed with 11 per subgroup.

The mean numbers of correct responses per *S* given to the even items during serial learning, arranged in serial position, may be summarized as follows: 36.36, 27.09, 25.36, 19.59, 22.55, 24.68, and 31.77. These even items were the response members of the seven experimental pairs in the subsequent PA learning. If it can be assumed that number of correct responses is a measure of associative strength, then the present experiment permits the formation of associations between sucessive items in the serial list.

The mean trials to learn the experimental and control PA sublists to one errorless recitation were 8.86 and 8.14, respectively ($F = .85$, $df = 1/20$). The mean correct responses per *S* during learning for the experimental and control sublists were 49.59 and 51.91, respectively ($F = .84$, $df = 1/20$). As in Exp. I, transfer of serial position effects, transfer during early learning, and overt errors were analyzed. In none of these analyses was a significant difference obtained.

EXPERIMENT II

Method.—The only difference between Exp. II and Exp. I was in the construction of the PA list. Two 15-item serial lists (plus a cue symbol) were constructed and a 10-item PA list was constructed from the two serial lists. Each stimulus in the PA list was composed of two items. These items were arranged one over the other and were in the same sequence (from top to bottom) as they were in the serial list. Forty-six *S*s served in the experiment with 23 learning each serial list before learning the PA list.

Results.—As in Exp. I, the PA list may be thought of as being composed of two five-pair sublists. The mean numbers of trials required to reach one errorless recitation of the experimental and control sublists were 12.04 and 9.63, respectively ($F = 8.35$, $df = 1/44$, $P < .01$). As can be seen from the data on trials to successive criteria presented in Fig. 3, negative transfer obtained from the serial to the PA list.

The mean numbers of correct re-

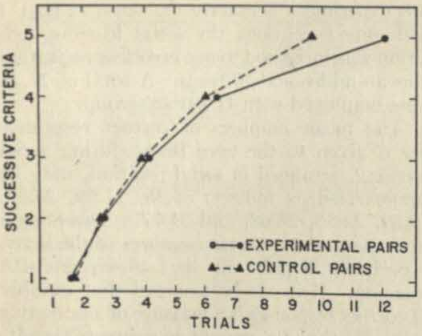


FIG. 3. Trials to successive criteria curves for the experimental and control sublists during PA learning in Exp. II.

sponses per *S* given during learning of the experimental and control sublists were 41.13 and 46.91, respectively ($F=7.27$, $df=1/44$, $P<.01$). Both response measures are in agreement that prior serial learning retards subsequent PA learning.

As in Exp. I, the pairs in the two PA sublists were arranged according to position in their respective serial lists and the mean percentage frequencies correct for the experimental and control pairs are presented in Fig. 4. Only one of the five comparisons was significant. The second experimental pair was more difficult than was the comparable control pair ($F = 20.49$, $df = 1/44$, $P < .001$).

The results of the error analysis in Exp. II are comparable to the error data obtained in Exp. I. Experimental stimuli elicit more overt errors than do control stimuli ($F = 19.70$, $df = 1/44$, $P < .001$) with mean errors per trial being .61 and .34 for the experimental and control sublists, respectively. In addition, more experimental items are elicited incorrectly by experimental stimuli (.56 per trial) than is the case for control stimuli eliciting control items (.29 per trial) ($F = 23.91$, $P < .001$). Again it should be noted that the remainder of the errors per trial, .05 and .05, is small and constant across conditions.

EXPERIMENT III

Method.—Experiment III differs from Exp. I and II in that transfer from one serial list to a second was studied and, in addition, a control group was employed. Four 13-item serial lists were constructed (with the first item serving as a cue symbol) for use in the experiment. Lists A and B were experimental lists and List C was a control list. Each *S* learned one of these lists and then all *Ss* learned the test list. Comparing List A with the test list, the even items in List A held the same serial position as they did in the test list while the odd items were randomly rearranged from List A to the test list. In List B the odd items held the same serial position as they did in the test list while the even items were randomly rearranged. In this manner test-list differences between those items with the same serial positions (S items) and those items with different serial positions (D items) are attributable to transfer effects rather than to differential difficulty due either to specific items or to serial position. The Control List C had no items in common with the test list, and test-list learning in the two experimental conditions may be compared to that in the control condition to evaluate transfer effects. Sixty-

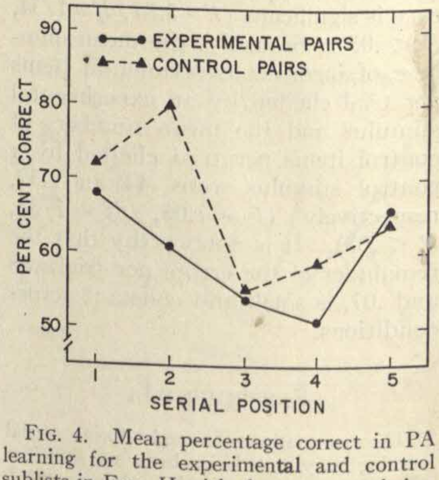


FIG. 4. Mean percentage correct in PA learning for the experimental and control sublists in Exp. II with the responsess being ordered along the baseline as they were in the serial lists.

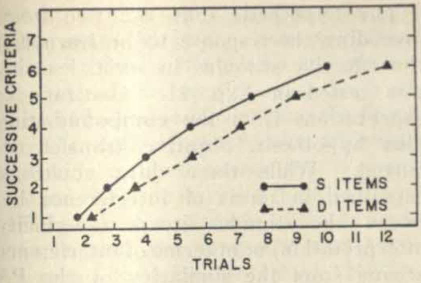


FIG. 5. Trials to successive criteria curves for the S- and D-item sublists during test-list learning in Exp. III.

three *S*s served in the experiment, 21 *S*s in each condition.

Results.—The mean numbers of trials to learn the test list in Conditions A, B, and C were 12.43, 13.86, and 9.71, respectively ($F = 6.50$, $df = 2/60$, $P < .01$). Orthogonal comparisons (Senders, 1958) between the three conditions indicate that the experimental conditions did not differ ($F = 1.50$) while the test list in the control condition was learned faster than in the experimental conditions ($F = 11.51$, $df = 1/60$, $P < .01$). Negative transfer obtained from the experimental to the test lists and these data are comparable to transfer data obtained by Irion (1946).

The test list may be thought of as being composed of two six-item sublists—the S items and the D items. Trials to reach a criterion of one errorless recitation for each of these sublists in the experimental conditions were analyzed. The mean trials to learn the S-item sublist was 10.17, while the mean for the D-item sublist was 12.14 ($F = 14.97$, $df = 1/40$, $P < .001$). The mean numbers of trials to attain successive criteria are plotted in Fig. 5.

The mean numbers of correct responses per *S* during test-list learning for the S- and D-item sublists were 46.86 and 39.90, respectively ($F = 35.24$, $df = 1/40$, $P < .001$). Both response measures are in agreement that the S-item sublist was easier to learn than the D-item sublist.

The S and D items were arranged according to their respective serial positions in the test list and, as can be seen from Fig. 6, differences obtained as a function of serial position. Only the first and last items did not differ between the S-item and D-item sublists. The *F*s for Serial Positions 2, 3, 4, and 5 ($df = 1/40$ in each case) were 6.82 ($P = .05$), 6.54 ($P = .05$), 10.59 ($P = .01$), and 6.79 ($P = .05$), respectively. Thus the items in the middle of the list were learned faster when the serial position was retained than when it was changed while such differences were not observed for the items at the ends of the list.

DISCUSSION

Experiment I tested the hypothesis that associations are formed between successive items within a serial list. This hypothesis, called the specificity hypothesis, was tested by constructing pairs taken from successive items in a serial list and measuring transfer from the serial list to a PA list composed of

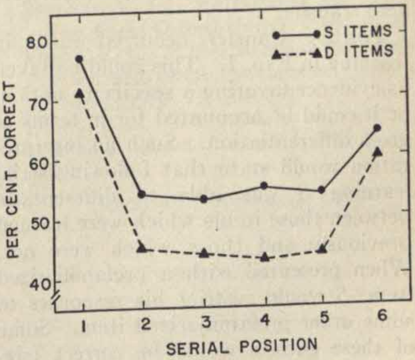


FIG. 6. Mean percentage correct in test-list learning for the S- and D-item sublists with the items being ordered along the baseline as they were in the test list.

the serial pairs. The analyses of the data indicated that, aside from positive transfer early in learning, the experimental pairs were learned no faster than the control pairs.

It may be argued that the absence of an overall difference between the two sublists was due to interference resulting from the change of tasks. However, when the arrangement of the items between serial and PA task is such as to expect negative transfer, zero transfer also occurs (Young, 1961).

On the other hand, the serial-position hypothesis assumes that the position the item holds in the list is the effective stimulus rather than the item immediately preceding the response to be learned. If such were the case, no transfer would be expected from the serial to the PA list because the stimulus in the PA learning would be different from the effective stimulus in serial learning. Thus, according to the serial-position hypothesis, the transfer paradigm employed in Exp. I would be an A-B, C-B paradigm of zero transfer. In this case the A stimulus would correspond to the serial position of the item in the serial list, the C stimulus would correspond to the nominal stimulus in the PA list, and the B response would be the response to be learned in both lists. Thus, while the specificity hypothesis predicts positive transfer from the serial to the PA list, the serial-position hypothesis predicts zero transfer.

Positive transfer occurred early in learning in Exp. I. This could be taken as evidence favoring a specificity notion, or it could be accounted for in terms of gross differentiation. Such an interpretation would state that following serial learning *S* was able to differentiate between those items which were learned previously and those which were not. When presented with a prefamiliarized item, *S* would restrict his responses to some other prefamiliarized item. Some of these guesses would be correct (significantly more than the control early in learning) and some would be incorrect (significantly more than the control errors).

The hypothesis that the two items preceding the response to be learned is the effective stimulus in serial learning was tested in Exp. II. Contrary to expectations from the compound-stimulus hypothesis, negative transfer occurred. While these data could be explained in terms of interference between the stimulus items (specificity interpretation) or in terms of interference arising from the similarity of the PA stimulus items to the initial serial positions (serial-position interpretation), the results of Exp. II do not permit serious consideration of the compound-stimulus hypothesis as a useful explanation of the stimulus in serial learning at this time.

The hypothesis tested in Exp. III was that the position the item holds in the serial list is the effective stimulus in serial learning. This hypothesis was tested by observing transfer from one serial list to another. The same items were employed in both lists and were arranged so that S items retained the same serial position in both lists while D items were randomly rearranged from the first to the second list. If the serial position were the effective stimulus, positive transfer would be expected for the S items and negative transfer would be expected for the D items. The results of the experiment show that the S items were easier to learn than the D items and that the test list in the experimental conditions was learned slower than in the control. The specificity hypothesis would make the prediction that no differences would obtain between the S and D items in the test list. According to this position, each item was associated to a response in the first list; but in no case were the same two items learned successively in both lists. Thus, the transfer paradigm for all items becomes an A-B, A-Br paradigm of negative transfer.

The addition of a mediation construct would permit interpretation of the results of Exp. III in terms of the specificity hypothesis. The only requirement would be the assumption that during test-list learning an S item acts as a stimulus and elicits the next (missing) item, which in turn correctly elicits the next (S) item. Such a mediation assumption would enable the specificity hypothesis to predict that the S items would be learned faster than the D items.

The data of the present experiments provide no final answer to the question of the effective stimulus in serial learning. They are, for example, difficult to integrate into the extensive literature on remote associations (e.g., McGeoch, 1936) which may best be interpreted in terms of the specificity hypothesis. However, the serial position curves of Exp. I and III (Fig. 2 and 6) may be used to formulate a tentative hypothesis which employs both the specificity and the serial-position hypotheses. Evidence from Exp. I and elsewhere (Young, 1961) indicates that learning in accordance with the specificity hypothesis occurs at the extremes of the serial list but not in the middle. In addition, data from Exp. III indicate that position learning has an important cue function in the middle of the list but not at the extremes. It then may be that specificity learning occurs at the extremes and that position learning occurs in the middle of the list. Some difficulties are immediately apparent with this formulation: the middle of the list is the point of greatest "position confusion" (Schulz, 1955), i.e., position learning should be less likely to occur; and when degree of serial learning is increased as in Exp. Ia, evidence for specificity learning is not apparent.

SUMMARY

Three hypotheses regarding the effective stimulus in serial learning were tested. If associations are formed between items of a serial list, then these associations should transfer to a subsequent PA list composed of the serial items with the S-R arrangement of the PA list being consistent with the ordering of the items in the serial learning. Positive transfer was observed only early in learning. If the effective stimulus in serial learning is the two preceding items in the list, then when sets of two items are used as stimuli, positive transfer should be observed in a subsequent PA task. In this case, negative transfer occurred. Finally, if the position of the item in a serial list is the effective stimulus, positive transfer should obtain from one serial list to another when the item retains the same serial position; and negative transfer should obtain when the serial position is changed. It was found that items retaining the same serial position were learned faster than items which had their serial positions changed from one list to another.

REFERENCES

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 517–567.

IRION, A. L. Retroactive inhibition as a function of the relative serial position of the original and interpolated items. *J. exp. Psychol.*, 1946, **36**, 262–270.

MCGEOCH, J. A. The direction and extent of intra-serial associations at recall. *Amer. J. Psychol.*, 1936, **48**, 221–245.

PORTER, L. W., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

SCHULZ, R. W. Generalization of serial position in rote serial learning. *J. exp. Psychol.*, 1955, **49**, 267–272.

SENDERS, V. L. *Measurement and statistics.* New York: Oxford Univer. Press, 1958.

YOUNG, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

YOUNG, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, **74**, 517–528.

(Received December 24, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 314–319

VERBAL CONCEPT ATTAINMENT: A FUNCTION OF THE NUMBER OF POSITIVE AND NEGATIVE INSTANCES PRESENTED

M. S. MAYZNER

University of Southern California

Kendler and his associates have suggested in a number of studies (Kendler & D'Amato, 1955; Kendler & Mayzner, 1956; Kendler & Vineberg, 1954) that the problem of concept attainment may be analyzed in terms of an implicit mediational mechanism. Most recently Kendler and Karasik (1958) have applied this approach to verbal concept attainment, and in the present study, we shall be concerned with developing this approach further.

In verbal concept attainment *S* typically is confronted with an array of instances or words. Usually, some subset of the total array is made up of those instances that *E* has designated as positive concept instances (i.e., words that belong to the concept), while the remaining subset is made up of those instances that *E* has designated as negative concept instances (i.e., words that do not belong to the concept). The *S*'s task is to identify or separate the positive from the negative instances, and to respond to the positive instances with the correct concept name.

According to a mediational hypothesis, *S* learns or attains the correct concept name in the following way. Each instance or word, whether positive or negative, evokes an implicit response. When more than one instance evokes the same implicit response, the probability of *S* grouping these different instances together increases, resulting in *S* explicitly responding to these different instances or words with the same concept name. If the instances selected and the associated concept corresponds to that subset of the total array that *E* designated as positive concept instances, then *S* is said to have attained the concept.

The employment of such a mediational model suggests a number of hypotheses concerning verbal concept behavior. First, it suggests that the ease or speed with which a given concept is attained is a function of the response strength of particular instances to evoke particular implicit responses. This hypothesis has been tested by Underwood and Richardson (1956b), who showed that the higher the dominance level (i.e., response strength), the more rapidly the concept was attained. Secondly, since implicit responses are evoked by all the instances, whether positive or negative, the type of implicit responses evoked by the negative instances (i.e., whether they are all different, all the same, have high or low response strength, etc.) may be as critically important as the implicit responses evoked by the positive instances. In fact, Kendler and Karasik (1958) have shown that different sets of negative instances, varying in the type of implicit responses they evoke, significantly affect the attainment of concepts based on a given set of positive instances. Thirdly, it might be expected that the actual number of positive and negative instances occurring in the total array should have a significant effect on concept attainment; that is, as the

TABLE 1

Concepts, Associated Positive Instances, and All Negative Instances Used in the Study

Test Problems

Concepts and Positive Instances								Negative Instances			
Round	%	Small	%	White	%	Hard	%				
BRACELET	25	POLLEN	30	HOSPITAL	32	ARMOR	24	ALUMINUM	CINNAMON	GINGER	OCEAN
MOON	30	THIMBLE	37	FROST	34	ASPHALT	29	APE	CLOVE	GYM	PINE
DIME	30	CABIN	39	BREAD	35	HELMET	31	BUTTER	COFFEE	HOG	SEAWEED
DERBY	33	SNAIL	42	PEARL	37	BRICK	35	CAMEL	CUSTARD	HONEY	SILK
PLATTER	38	WAIST	43	COLLAR	44	SKULL	36	CHEESE	FLANNEL	MOCCASIN	TAR
								CHOCOLATE	FOREST	MOSS	TOBACCO

Practice Problems

Concepts and Associated Positive and Negative Instances					
Red	%	Negative Instances	Sharp	%	Negative Instances
CHERRY	77	AMMONIA	FISHHOOK	70	CITY
TOMATO	83	BED	DAGGER	75	CORN
BEET	87	DANDELION	HATCHET	77	GARLIC
BLOOD	91	ELEPHANT	SCISSORS	78	GRASS
BLUSH	96	IVY	KNIFE	84	NIGHT
		VINEGAR			VELVET

number of positive instances, at a given level of response strength, is increased, there should occur a corresponding increase in the probability of more than one of them evoking the same implicit response, which should produce more rapid concept attainment. Conversely, as the number of negative instances is increased, when each negative instance is designed to evoke a different implicit response, there is a corresponding increase in the total number of different implicit responses evoked, which might be expected to interfere somewhat with the evocation of a single implicit response to the positive instances, thus producing less rapid concept attainment. It is to this issue, that is the effect of the number of positive and negative instances occurring, that the present study is directed.

METHOD

Subjects.—The *S*s were 144 students from introductory psychology classes at New York University, 12 in each of 12 different experimental conditions.

Stimulus material.—The stimulus material used in the study was selected from a word-list prepared by Underwood and Richardson (1956a) which gives the response strength (i.e., percentage frequency) of each word to evoke a sense-impression response. Since each *S* was required to attain four concepts, namely, Round (R), Small (S), White (W), and Hard (H), four lists of five positive instances were selected from Underwood's list, and are presented in Table 1, along with the percentage frequency value with which each instance evokes its associated concept.

As can be seen in Table 1, the range of response strengths associated with each concept was kept as low and as nearly equal, as was possible, to provide a moderately difficult and approximately equivalent set of problems, which might be expected to most clearly reveal the effects of the variables under investigation. Also included in Table 1 are the 24 negative instances employed in the study. These negative instances were selected from Underwood's list on the basis of two criteria: (*a*) none possess any response strength to evoke any of the four problem concepts, at least as indicated by Underwood's norms; (*b*) the response strength to evoke other concepts was kept as low as possible, in order to examine at a base level the effect of increased numbers of negative instances.

Procedure.—The present study is concerned with examining the effects of the number of positive and negative instances occurring, on the speed of concept attainment. To examine these effects, a 4 × 3 factorial analysis of variance design was selected. Thus, either two, three, four, or five positive instances were combined with either two, four, or six negative instances to yield 12 different experimental conditions, with 12 *S*s per condition. In order to minimize the possibility that the results might be a function of a specific concept problem, each *S* received four test problems involving the Concepts R, S, W, and H. Four different orders of presentation were employed, namely, RWSH, SRHW, WHRS, and HSWR, i.e., within each condition 3 *S*s received the order RWSH, another 3 *S*s received the order SRHW, etc.

The form in which each problem was presented was as follows. All instances, both positive and negative, were typed in lowercase letters on white 3 × 5 in. cards. Each problem array was placed on a table before *S*. At the start of each problem, *S* was told to look at the cards and *E* started a stopwatch. The problem array may best be understood from the instructions given to *S*, as follows:

> When I tell you to start the problem, you will notice before you two rows of cards with words typed on them. The top row, you will notice has four cards or words, while the bottom row has seven cards or words. There is a single word in the bottom row, which may be grouped with the four words in the top row, on the basis of a common sensory quality all five words share together. Your task is to point out which word in the bottom row may be grouped with the four words in the top row, and to indicate the basis of the grouping, i.e., tell me the common sensory quality on which you have based your selection.

These instructions apply to that experimental condition involving five positive instances and six negative instances, with analogous instructions tailored to fit the other 11 conditions, with respect to the number of positive and negative instances occurring. Thus, one positive instance always appeared in the bottom row of words, with either two, four, or six negative instances, and either one, two, three, or four positive instances always appeared in the top row of words.

The *S*s were assigned at random to the 12 experimental conditions. Each *S* received two practice problems and four test problems and the time to solution was taken as the dependent variable for study. In the two practice problems, the number of positive and negative instances was the same as the number of positive and negative instances presented in the four test problems. The practice problems, shown in Table 1, were made easier than the test problems to insure that *S* fully understood the nature of the problems and what was required of him to achieve a correct solution. The *S* was instructed that a correct solution involved both (*a*) choosing the correct instance in the bottom row of words, and (*b*) applying the correct concept name. Either one alone was not considered a correct solution and *S* was instructed to continue working on the problem when this occurred. If no correct solution was obtained within 2 min., *S* was given the correct solution, and the next problem was presented.

Although care was taken to insure that all positive instances belonging to a given concept were essentially equivalent (i.e., had approximately equal percentage frequency values) additional design precautions were taken in the construction of each of the four problems presented to *S*, as follows. In all conditions involving the use of five positive instances (four in the top row and one in the bottom row) each of the five positive instances occurred at least twice and two of the five occurred three times, in the bottom row, to insure that no one particular positive instance appeared relatively more or less frequently than any other positive instance, among the negative instances. Similar precautions were taken with the conditions involving four, three, and two positive instances and, in addition, the selection of the remaining positive instances for the top row (either three, two, or one) was designed to insure relatively equal frequency of use. The combinations or arrays of positive instances that were so constructed, were employed without change in the conditions involving two, four, or six negative instances. Further, the placement of the positive instance among the negative instances was designed to insure that the positive instance occurred with approximately equal frequency in each of the available positions in the row.

Precautions also were taken in the selection of the negative instances. In the conditions involving 6 negative instances, since each *S* received 4 problems, the entire pool of 24 negative instances was utilized, as follows. The 24 negative instances were

TABLE 2

MEDIAN SOLUTION TIMES (SEC.) AND NUMBER OF SOLUTIONS FOR EACH PROBLEM-POSITION AND EACH CONCEPT SUMMED ACROSS THE 12 EXPERIMENTAL CONDITIONS

Concepts	Measure	Problem-Position				
		1	2	3	4	Totals
Round	Median	35.0	25.0	18.5	20.5	22.0
	No. solving	24	24	27	27	102
Small	Median	120.0[a]	120.0	120.0	120.0	120.0
	No. solving	7	7	13	11	38
White	Median	77.5	57.0	25.5	56.5	56.0
	No. solving	21	21	26	22	90
Hard	Median	57.0	19.5	70.0	22.5	32.0
	No. solving	22	30	23	28	103
Totals	Median	97.0	62.0	41.5	41.5	
	No. solving	74	82	89	88	

Note.—$N = 36$ each cell.
[a] Time-limit = 120 sec.

typed on cards and shuffled, 6 cards were then drawn at random, then from the remaining pack of 18 cards, 6 more were drawn at random, and the remaining pack of 12 cards was then randomly separated into 2 packs of 6 cards each. The 4 separate packs were then randomly assigned to serve as the 6 negative instances, for each of the 4 problems presented to a given *S*. This procedure was repeated 12 times, to provide a different random order of negative instances for each of the 12 *S*s in a given experimental condition. The same 12 random orders were employed, however, in the conditions involving 2, 3, 4, and 5 positive instances. Analogous randomization procedures were applied to the conditions with 2 and 4 negative instances, but with only 2 or 4 cards being drawn at each drawing, in order to construct the separate problems. When the separate drawings, of either 2 or 4 cards, exhausted the total pool of 24 cards, the 24 cards were again shuffled and the drawing procedure was repeated.

RESULTS

The dependent variable selected for study was "time to solution" and since each *S* received four problems, it was intended to take the median of these four time scores for each *S*, as the measure for analysis of the main effects in the design. However, the problems were not of equal difficulty as shown by the results presented in Table 2, which gives the median solution times and the number of solutions attained within the 120-sec. time limit, for each problem-position and each concept summed across the 12 experimental conditions. These results indicate that the concept Small is a far more difficult problem than the other three concepts, which are approximately of equal difficulty. This conclusion is supported by a chi square analysis ($\chi^2 = 34.52$; $df = 3$; $P < .001$) contrasting the obtained number of solutions for all four concepts (R-102, S-38, W-90, and H-103) with the expected equal number. Further, there is no significant difference between Concepts R, W, and H ($\chi^2 = 1.07; df = 3; P > .70$). The effect of problem-position is clearest in the Totals column and the correlated *t* for the difference between Problem-Positions 1 and 4 was 2.15 ($df = 143$; $P < .05$), suggesting a significant practice effect for this type of problem.

Since the concept Small produced such a deviant distribution of time scores, the median of the three time scores on Concepts R, W, and H was selected as the criterion measure for use in the analysis of variance. Thus, there was made available for each *S* a single median time score, and means, *SD*s, and ranges of these median scores, for the 12 experimental conditions of the study, are presented in Table 3. The results in Table 3 indicate that, in general, concept attainment solution times increase with increased numbers of negative instances, but decrease with increased numbers of positive instances, as most clearly seen in the row and column means. An analysis of variance yielded *F* values of 6.33, 1.40, and .39, for the main effects associated with the number of positive instances, the number of negative instances, and the interaction term, respectively, with only the *F* associated with the number of positive instances reaching statistical significance ($df = 3/132$; $P < .001$). Bartlett's test yielded a chi square of 8.60 ($df = 11; P > .50$), indicating the analysis did not violate the assumption of homogeneity of variance.

Discussion

It was predicted, on the basis of an S-R mediational model, that increasing the number of positive instances should produce more rapid concept attainment, since each additional positive instance presented to *S* should increase the probability of more than one such positive instance evoking the same implicit response, and matching what *E* has designated as the c[illegible]t concept. It also was predicted, th[illegible]asing the number of negative in[illegible]en each negative instance is designed to evoke a different implicit response, should produce less rapid concept attainment, since the presence of many different implicit re-

TABLE 3

Concept Attainment Solution Times (Sec.) for Different Numbers of Positive and Negative Instances

Number of Negative Instances		Number of Positive Instances				
		2	3	4	5	Row Mean
2	Mean	68.92	57.00	38.25	30.25	48.61
	SD	51.59	39.71	39.58	31.83	
	Range	10–120	9–120	6–120	9–120	
4	Mean	77.08	42.08	57.17	36.75	53.27
	SD	42.30	37.27	46.56	24.92	
	Range	5–120	5–115	6–120	9–99	
6	Mean	90.00	57.25	62.42	41.25	62.73
	SD	32.08	46.30	43.54	39.53	
	Range	30–120	8–120	7–120	5–120	
Column Mean		78.66	52.11	52.61	36.08	

Note.—Based on medians for individual *S*s.

sponses might be expected to interfere with the evocation of a single implicit response to the positive instances. The results indicate that while the effects of both positive and negative instances are in the predicted direction, only the effect associated with the positive instances reaches statistical significance.

The failure of the negative instances to produce a significant effect might indicate that the negative instances do, in fact, have no effect on concept attainment. However, since three of the four columns in Table 3 (i.e., two, four, and five positive instances) show systematic increases in solution times as the number of negative instances are increased from two to six, it seems more reasonable to conclude that the negative instances produce the effect predicted, but that the effect is a weak one and the present design was not sufficiently sensitive to measure it.

One might ask under what conditions would an increase in negative instances be expected to produce a strong effect. In a mediational model, three principal factors determine the speed with which a concept is attained. They are: (*a*) the number of positive instances presented and the response strength associated with each positive instance to evoke the appropriate implicit response; (*b*) the number of negative instances presented and the response strength associated with each negative instance to evoke inappropriate or irrelevant implicit responses; and (*c*) any interaction which occurs between the implicit responses evoked by all the instances, both positive and negative (i.e., Kendler and Karasik's (1958) "response differentiation"). By systematically varying then the parameters of this model, one could determine the particular values or conditions necessary for the negative instances to produce a strong effect.

SUMMARY

The present study examined the effects of the number of positive and negative instances on verbal concept attainment. Four values of positive instances (two, three, four, and five) were combined with three values of negative instances (two, four, and six) in a 4 × 3 factorial design with 12 *S*s per condition. The results, predicted on the basis of an S-R mediational model, showed that increased numbers of positive instances produced significantly more rapid concept attainment, while increased numbers of negative instances produced, though not significantly so, less rapid concept attainment.

REFERENCES

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

KENDLER, H. H., & KARASIK, A. D. Concept formation as a function of competition between response produced cues. *J. exp. Psychol.*, 1958, **55**, 278–283.

KENDLER, H. H., & MAYZNER, M. S. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

KENDLER, H. H., & VINEBERG, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, **48**, 252–258.

UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95. (a)

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229–238. (b)

(Received December 30, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 3, 320

SUPPLEMENTARY REPORT: MONETARY INCENTIVE AND MOTIVATION IN DISCRIMINATION LEARNING—SEX DIFFERENCES

BETSY WORTH ESTES, LOUISE BRIGHTWELL MILLER,[1] AND MARY ELLEN CURTIN
University of Kentucky

Miller and Estes (1961) found that amount of monetary reward, even in a 50 to 1 ratio, failed to improve error scores in the acquisition of a visual discrimination by 9-yr-old boys. A detrimental reward effect was suggested, in that rewarded groups made significantly more errors than a knowledge-only (KOR) group. n Achievement (n Ach) was shown to be a significant variable, high scorers making fewer errors. This study provides similar comparisons for adult *S*s.

Method.—The *S*s were 48 female and 72 male students in introductory psychology classes. A 3 × 2 × 2 factorial design was used: (*a*) three incentive levels (knowledge of results [KOR], 1¢, and 25¢); (*b*) two n Ach levels (high and low); and (*c*) two sex levels. Design, apparatus, and procedure were the same as in the previous experiment, except that the high reward group received 25¢ per correct choice instead of 50¢, and the stimulus presentation interval was shortened to 1 sec. in order to increase the difficulty of the task for adults. Groups were matched on the basis of n Ach scores, father's occupation, and amount of financial support from parents.

Results.—The significant variables were sex (females made fewer errors) and the sex-incentive interaction, Table 1. Sex differences and interaction are shown in Fig. 1. The detrimental effect of the 1¢ reward for females is confirmed by *t* tests; there was no significant incentive effect among males.

Female superiority might reasonably be attributed to cultural differences between sexes in past experience and training since the task involved attention to minute facial detail. Although a detrimental effect with low reward was also found in the previous study, the parallel between the 9-yr-old boys and adult females ends here. With 50¢ reward, performance of the young boys was similar to that of the 1¢ group and hence was poorer than with KOR; performance of the adult females with higher reward (25¢) was *better* than that of the 1¢ group and similar to KOR.

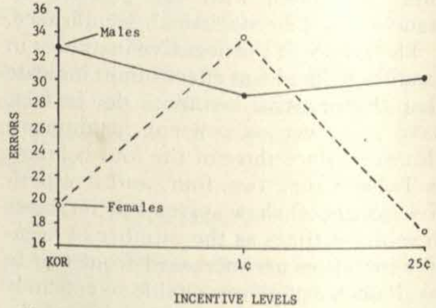


FIG. 1. Male and female performance according to incentive levels.

TABLE 1

ANALYSIS OF VARIANCE OF ERRORS

Source	*df*	*MS*	*F*
Sex (S)	1	1,433.69	5.30*
n Ach (N)	1	444.68	1.64
Incentive (I)	2	334.30	1.24
S × N	1	55.20	.20
S × I	2	1,025.17	3.79**
N × I	2	211.60	.78
S × I × N	2	444.32	1.64
Within cells	108	270.69	
Total	119		

* $P < .025$.
** $P < .05$.

Age differences are shown not only by lack of incentive effect for adult males, but also by the n Ach factor, significant for young boys but not for male or female adults. One tenable explanation is that n Ach in adults was not measured by the method used. Another explanation involves the problem of reproduction of instructions when repeating with adults studies done with children. Atkinson (1954, p. 75) reports that high n Ach *S*s perform better "only when the instructions defined performance as instrumental to producing feelings of accomplishment." Although instructions were experimentally achievement-oriented, it seems unlikely that the task of learning to tell drawings of twins apart could offer as much satisfaction in accomplishment for college students as for elementary school children, regardless of instructions.

REFERENCES

ATKINSON, J. W. Exploration using imaginative thought to assess the strength of human motives. *Nebraska symposium on Motivation: 1954.* Lincoln: Univer. Nebraska Press, 1954.

MILLER, L. B., & ESTES, B. W. Monetary reward and motivation in discrimination learning. *J. exp. Psychol.*, 1961, **61**, 501–504.

(Received March 23, 1961)

---

[1] Now at the University of Louisville.

Journal of
Experimental Psychology

VOL. 63, NO. 4 APRIL 1962

COGNITIVE ASPECTS OF INFORMATION PROCESSING: I. THE FAMILIARITY OF S-R SETS AND SUBSETS[1]

PAUL M. FITTS AND GAIL SWITZER

University of Michigan

Choice reaction time (RT) has been shown to be a function of several classes of variables, including discriminability (Crossman, 1955), S-R compatibility (Fitts & Seegar, 1953; Morin & Grant, 1955; Welford, 1958, Ch. 6), and amount of information to be transmitted. The present series of studies was designed to investigate the effects of certain cognitive variables, which can be viewed as influencing Welford's (1960) translational stage of information processing. In particular, the present studies deal with *S*'s ability to adopt and utilize cognitive sets appropriate to the number of alternatives permitted to occur in an experiment. The use of familiar vs. unfamiliar groups of stimuli, *considered as a set*, is emphasized, in contrast to the usual emphasis on the familiarity, or response strengths, of individual S-R associations.

Mental set, expectancy, and related concepts have, of course, been frequently invoked in psychology for many years (see reviews by Dashiell, 1940; Gibson, 1941). For example, Helson (1959) has emphasized the widespread importance of adaptation level as influenced by background and contextual factors, Postman and Bruner (1949) have demonstrated the greater efficiency of perception under single than under multiple sets, and Harlow (1949) has developed the concept of learning sets. However, the importance of cognitive sets in influencing the efficiency of performance in information handling (speeded perceptual motor) tasks has not been systematically investigated.

The importance of cognitive set as a factor influencing information handling efficiency, is indicated by an examination of some apparently conflicting results which have appeared recently from experimental studies of choice RT. One of the most generally accepted empirical laws governing perceptual-motor behavior is the one relating RT to stimulus uncertainty, which has been expressed by some writers as:

$$\mathrm{RT} = A + BH_t \qquad [1]$$

H_t is average information transmitted per response.

A and B are constants.

This linear increase in RT with increasing information transmitted, first emphasized by Miller (1951), Hick (1952), and Hyman (1953), has been verified by numerous subsequent experiments. A few studies, however, have revealed instances in which RT is essentially independent of information, i.e., in which the slope constant, B, in Equation 1

---

[1] This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command under Contract No. AF 49 (638)-449.

approaches a value of zero. Conditions under which the latter relation has been found appear to be characterized by the use of particular stimulus and response codes, and/or by highly compatible or overlearned S-R sets.

When the response required of *S* is to move one of his fingers, RT usually has been found to increase as H_t increases, over the range of 1 to 10 fingers. However, once *S* is required to use 2 or more fingers in combination, results (Klemmer & Muller, 1953; Seibel, 1960) indicate that variations in the size of the subset of finger combinations used in the task has little or no effect on RT. These effects have been found after small or moderate amounts of training; very extensive training lasting over many weeks may, of course, give different results.

Other instances of very small changes in RT as a function of stimulus uncertainty have been reported when responses were made to English language symbols (words, letters, or numerals). Quastler and Brabb (1956) found very small changes in typing speed when copy was drawn from alphabets of differing size. Similarly, Pierce and Karlin (1957) found that alphabet size per se, over the range from 2 to 256 words, had little effect upon speed of reading lists of random words aloud, provided word length and frequency of general usage in the English language (i.e., familiarity) were held constant. Mowbray (1960) found no difference in RT in naming numbers, over the range from 2 to 10 alternatives. Davis, Moray, and Treisman (1961) played numerals, letters, and words from a tape recording into earphones and required *S*s to speak each sound as soon as possible after hearing it. Responses to numerals were slightly faster (by about 20 msec.) than to the larger alphabet of letters, but no effects of varying the size of the subsets of numerals or of letters were found.

The dependence of the constants in Equation 1 on coding variables is further shown in a study by Brainard, Irby, Fitts, and Alluisi (1962) employing four different S-R ensembles. In three of their four S-R codes a reduction in RT followed a reduction in number of alternatives. In the fourth, however, involving vocal (naming) responses to numerals, no significant changes in RT were found for two, four, or eight alternatives. The authors suggest that this lack of an effect may have been due to the manner in which stimulus uncertainty was reduced. In the two-alternative case, for example, 4 and 7 were used as stimuli and it was noted that this pair does not form a familiar subset. Newman (1959) has offered a similar hypothesis, namely that ". . . short-term instruction is probably not effective in reducing the complexity of the task in the face of a history of many hours of practice . . ." (p. 14).

Leonard (1959), using tactual stimuli applied to the finger tips, and Mowbray and Rhoades (1959), training 1 *S* for several months with lights as stimuli, both found little or no effect of number of alternative finger responses even when responses were made with single fingers. Leonard (1961) hypothesizes that number of alternatives has the least effect when compatibility, discriminability, and/or familiarity are high.

In summary, all of the exceptions to the general finding that RT increases with stimulus uncertainty appear to indicate the importance of coding variables, or, more specifically, the importance of (*a*) the natural or learned relations between individual stimuli and responses, and (*b*) familiarity in dealing with particular sets and subsets of stimuli and responses. The outline of a general theory relating to the latter type of effect can be stated as follows.

First, it appears that facility in dealing with a particular ensemble of stimuli develops slowly and for the most part continuously with practice. Conceptual understanding of the requirements of a task may be rapid, but speed and accuracy of responses will usually continue to improve slowly, over a period often measured in years, depending on the complexity of the situation (alphabet size, laws of syntax, etc.). Thus, speed and accuracy of choice in dealing with a particular event is assumed to depend in part on the amount of previous practice in using the cognitive set appropriate to the over-all task or situation in the context of which the event occurs.

Second, it is assumed that a cognitive set, once it has been thoroughly established, can become dominant in as short a time as a second or less, as a result of the occurrence of an appropriate cue. Since a particular S-R pair may be included in several cognitive sets of differing size, the speed and accuracy of the response to any given stimulus will be in part a function of the set which is dominant at the time the response is made as well as of S-R association strength per se.

Third, it is assumed that, when *S* is asked to perform a new task, some specific or multiple cognitive set will become dominant. This initial cognitive set may include excess or inappropriate stimuli and responses, and some appropriate stimuli and responses may not be included. In the words of Postman and Bruner (1949) it may not be "finely tuned." Such a cognitive set for a given task can be thought of as analogous to a computer program, governing a data processing task, such as the carrying out of a search procedure or the application of a statistical decision rule, and an associated set of data in a rapid-access memory; an inappropriate cognitive set is then analogous to a program and a set of data not entirely appropriate to the task at hand (i.e., to the type of decision or search required).

The present experiments were planned to test certain of these assumptions, taking advantage of *S*s' extended experience with sets of English alpha-numeric symbols. Prior to each experimental session *S*s were always told exactly what stimuli to expect and what responses to make during a session. However, it was postulated that they would be able to prepare adequately for the specific task, in advance, only if they had previously developed a cognitive set appropriate to the specific ensemble of stimuli and responses to be used in the task.

METHOD

Apparatus

Stimuli were presented by means of a projection-type "in-line" display (Exp. 1) or a double-bladed shutter which exposed white cards carrying black letters (Exp. 2 and 3). The *S*s sat in front of a large gray screen and stimuli appeared in a center window. They wore Wilson "sound barrier" headsets and listened to broad-band masking noise except when receiving communications from *E*. Onset time of vocal responses was measured by use of a boom microphone, electronic voice key, and .01-sec. time clocks. The lag time of the apparatus (shutter, voice key, relays, etc.) was .04 sec.; this lag was subtracted from *S*s' response times.

Subjects

The *S*s were 78 volunteer, university students who served for pay. In Exp. 1 and 3 three groups of 8 *S*s and three groups of 15 *S*s, respectively, including men and women, were used. In Exp. 2 one group of 9 *S*s, all men, was used.

Procedure

Experiment 1.—The 24 *S*s were pretested on a two-choice color-naming task and mean RT scores were used to form three strata; equal numbers were then assigned randomly from within each stratum to each of the three experimental groups, with the restriction that each group contain the same proportions of men and women.

Group 1-LF used the larger, familiar set of eight numerals, 1 through 8. Group 1-SF used a small familiar subset of two numerals, 1 and 2. Group 1-SU used a small unfamiliar subset of two numerals, 2 and 7 (2 was chosen to provide a stimulus common to all three conditions; 7 was chosen because it had been used by Brainard et al., 1962).

During Sessions 1, 2, and 3 each group was given 80 trials. Stimuli were presented in a different randomized order to each *S*, with the restriction of equal frequencies of alternatives over each session. Stimuli were exposed exactly 2 sec. after a warning signal (a .5-sec. interruption of the noise in the headset) and there was approximately a 10-sec. interval between stimuli. Stimuli disappeared as soon as *S* began to speak. Following the initial instructions *E* communicated with *S* only to give knowledge of results (fastest and slowest times) after each cycle of 10 responses.

Experiment 2.—Each of 9 *S*s served for 4 days and was tested with each of the following sets of stimuli on each day: (*a*) Numerals 1 and 2; (*b*) Numerals 1, 2, 3, and 4; (*c*) Numerals 4 and 7; (*d*) Numerals 4, 5, 6, and 7; (*e*) Numerals 1 through 8. A total of 200 trials was given during each session, 40 under each of the five conditions. The sequence was randomized and different for each *S* within blocks of 40 trials and counterbalanced across sessions. The *S*s were told what the set of stimuli would be before each block of trials.

Experiment 3.—Pretests on the color-naming task were given and the 45 *S*s assigned to three groups by the same method as in Exp. 1.

Group 3-LF used all 26 capital letters of the English alphabet. Group 3-SF used a small familiar subset of three letters, A, B, and C. Group 3-SU used the small unfamiliar subset, E, B, P. The letter B, which was common to all three conditions, was chosen because it is a hard phoneme, well suited for detection by the voice key. The other two letters of the unfamiliar subset, E, and P, were chosen because they have frequencies of usage in written English as well as phoneme frequencies similar to the letters A and C. Discriminability, and interpair confusion frequencies in legibility studies, are approximately equal (Howell & Kraft, 1959).

Each group was given 80 trials on Sessions 1, 2, and 3. Letter sequences were randomized over sessions, each letter occurring a minimum of three times per session. Other procedures were the same as for Exp. 1.

RESULTS

Reaction times.—Speaking time (time from initiation to completion of the vocal response) was found in preliminary tests to be of short duration and relatively invariant with conditions, when single letters or numerals were used. Therefore only RT data (time to initiate the vocal response) were employed in these experiments. Mean RT scores for each group on each session were computed. All results are consistent with the hypothesized differences between familiar and unfamiliar subsets of stimuli.

The most pertinent data for Exp. 1 are those for the numeral 2, which was common to all three groups (see Fig. 1). On Session 1 the RT to 2, when 2 occurred as an alternate to 7 (Group 1-SU), was as slow as when it occurred as part of a set of eight numerals. But when it occurred as an alternate to 1 (Group 1-SF) the average RT was about .02 sec. less than when it was part of the larger set. Groups 1-SF and 1-SU both showed some learning during Sessions 2 and 3, but the improvement was small. Less learning occurred for Group 1-LF.

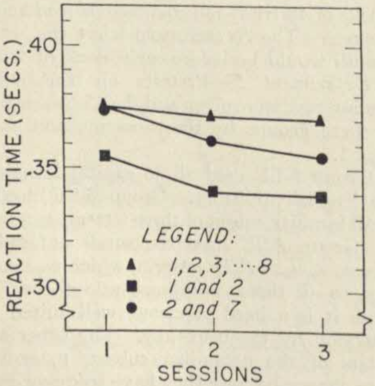


FIG. 1. Reaction times to the numeral 2 when it occurred as a member of three sets of numerals (Exp. 1).

TABLE 1

ANALYSIS OF RTs TO THE NUMERAL 2 (EXP. 1) WHEN CONTAINED WITHIN DIFFERENT SETS OF STIMULI

Source	*df*	*MS*	*F*
Between *S*s	23		
No. Alternatives (A)	2	47.37	2.49
Error (s)	21	18.96	
Within *S*s	48		
Sessions (D)	2	0.03	
A × D	4	1.47	2.88*
Error (w)	42	0.51	

* $P < .05$.

Results of an analysis of covariance for these data are given in Table 1 (covariance analysis was used instead of a Treatments × Levels analysis because the stratification employed in assigning *S*s to groups was quite coarse and did not make full use of the information in the pretests). Color-naming pretest scores correlated highly with test performance for all the three groups (r = .76, .74, and .82, respectively). The effect of Conditions was not significant, but the interaction of Sessions and Conditions was significant at the .05 level, indicating differential learning rates.

In Exp. 2, where *S*s also worked with numerals, but served as their own controls, the results were very similar. Data for familiar sets of stimuli containing the numerals 1 and 2 are summarized in Fig. 2A; results of an analysis of variance for these data are given in Table 2. For this familiar subset condition RT increased slightly, but significantly, for each increase in the number of equiprobable alternatives. In fact

TABLE 2

Analysis of Variance of RTs to Numerals When Contained within Sets of Differing Size and Familiarity

Source	*df*	Numerals 1 and 2		Numerals 4 and 7	
		MS	*F*	*MS*	*F*
No. Alternatives (A)	2	38.32	12.0**	7.49	2.44
Sessions (D)	3	38.15	11.2**	34.97	5.02**
Subjects (S)	8	115.47		119.28	
A × S	16	3.19		3.07	
D × S	24	3.40		6.97	
A × D	6	2.63	1.43	3.41	1.87
A × D × S	48	1.84		1.82	

** $P < .01$.

8 out of the 9 *S*s gave four-session individual mean RTs which showed a progressive orderly increase as a function of stimulus uncertainty, as Equation 1 would predict. For the unfamiliar subsets, however, RT (for 4 and 7) had no consistent relation to the size of the stimulus set (see Fig. 2B). Not 1 of the 9 *S*s had individual mean RTs for 4 and 7 which increased in an orderly fashion with stimulus uncertainty as predicted by Equation 1. The over-all *F* test for the effect of number of alternatives was not significant (see Table 2). The group means on the last two sessions actually were fastest for the eight-alternative condition and the differences in RT to 1 and 2 vs. 4 and 7, two-choice task, were greater, in fact, at the end than at the beginning of practice. Thus the same *S*s who were consistently able to give faster RTs to a familiar set, as stimulus uncertainty was reduced, were unable to do so when an unfamiliar set was used. The significance of this finding is further indicated (see Fig. 2A and 2B) by the fact that average RTs for these contrasting pairs (1 and 2 vs. 4 and 7) were almost exactly equal when both

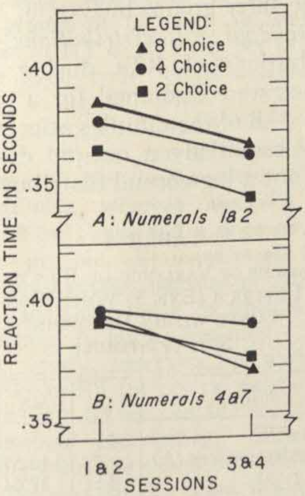


Fig. 2. Reaction times to several sets of numerals (Exp. 2). (In A are shown data for 1 and 2 when used alone and as a part of two other sets; in B are shown similar data for 4 and 7.)

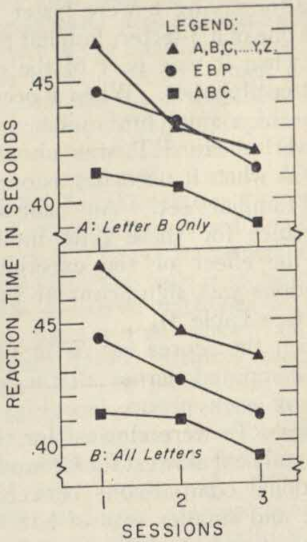


Fig. 3. Reaction times to three sets of letters (Exp. 3). (In A are shown data for the letter B when it occurred in each of these sets; in B are shown the means for all letters comprising the three sets.)

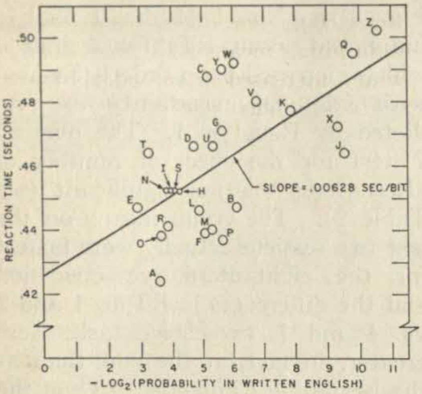


FIG. 4. Relation of English letter frequencies (Pratt, 1939) to vocal RT when all letters were used equally frequently in Exp. 3.

were contained in the same set of eight familiar numerals. Thus the differences were not due to association strength per se.

The results for Exp. 3 (see Fig. 3A), employing letters of the alphabet, are similar to those for Exp. 2. RT scores for saying B were faster when B was one of a 3-letter, familiar subset than when it was part of the entire 26-letter alphabet. When B occurred as one of a small unfamiliar subset, however, mean RT was about the same as when it occurred as one of a large familiar set. An analysis of covariance for these data indicated that the effect of the experimental conditions was significant at the .01 level (see Table 3).

When the scores for *S*s in Exp. 3 were computed across all the letters used by each group (see Fig. 3B), average RTs were slowest for Group 3-LF and next slowest for Group 3-SU. Additional comparisons between the larger and smaller sets of letters are also of interest. RTs to the letters A and C were 40 msec. faster when they occurred in the small familiar set than when they occurred in the large familiar set; however, RTs to E and P were only 3 msec. faster in the small unfamiliar set than in the large familiar set.

Correlations between relative letter frequencies in English (Pratt, 1939) and average RTs, were computed for Group 3-LF, and found to be moderately high (r = .60 for Session 1 and .62 for Session 3). The absolute magnitude of these RT differences, due to English letter frequencies, were of about the same magnitude as the maximum differences produced by reducing the size of the stimulus set in Exp. 3 (e.g., the difference between the most frequent letter (E) and the least frequent one (Z) was about 50 msec.). (See Fig. 4).

Errors.—The over-all percentages of errors in the three experiments were small (1.4%, 0.2%, and 0.6%) and no important differences in proportions of errors occurred under the different experimental conditions. Thus errors were 1.4%, 1.1%, and 1.6% for the three groups in Exp. 1, and 0.4%, 0.9%, and 0.4%, respectively, for the three groups in Exp. 3.

Individual RT distributions.—The distributions of RTs during single sessions were examined for a sample of *S*s. All of the samples appeared to follow essentially a normal distribution (not a log-normal distribution as

TABLE 3

ANALYSIS OF VARIANCE OF RTS TO THE LETTER B (EXP. 3) WHEN CONTAINED WITHIN DIFFERENT SETS OF STIMULI

Source	*df*	*MS*	*F*
Between *S*s	44		
No. Alternatives (A)	2	146.57	5.31**
Error (s)	42	27.61	
Within *S*s	90		
Sessions (D)	2	1.23	—
A × D	4	7.01	—
Error (w)	84	6.35	

** $P < .01$.

might have been expected). Practically all the mean RTs of all *S*s fell within a range of ±100 msec.

DISCUSSION

The results substantiate the prediction that a stimulus element, in a familiar subset, will elicit faster RTs than will the same element either when it occurs as part of a less familiar stimulus subset of the same size or as part of a large and equally familiar set. Demonstration of such effects requires careful control of differences in discriminability and association strengths of stimuli and responses; these are believed to have been adequate in the present experiments. Fully as important an aspect of the present findings, however, is the relatively small absolute magnitude of the obtained differences.

In some tasks relatively large increases in choice times result from increases in stimulus uncertainty. Hyman (1953), for example, using lights as stimuli for vocal responses and employing 4 highly practiced *S*s, found a median increase in RT of about .176 sec. per bit of stimulus uncertainty. Brainard et al. (1962), using groups of 12 naive *S*s, found an increase in RT of .178 sec. per bit for a similar S-R code (vocal-numeral responses to lights). Hick (1952), for 1 highly practiced *S*, found an increase of about .11 sec. per bit for finger responses to lights (in the range up to 10 alternatives), whereas Merkel's data, reported by Hick, had a slope of about .13 sec. per bit. Brainard et al. (1962) found an average increase of .136 sec. per bit for their naive *S*s when an S-R code similar to Hick's was used. In the present study, however, the maximum difference in RT for two vs. eight numerals was only .03 sec., or .015 sec. per bit of stimulus uncertainty. Differences resulting from the use of 3 vs. 26 letters were also very small. Thus the present results are similar to those reported by investigators who have studied responses to English language symbols. Even when we compare numerals with letters, two highly familiar alphabets of differing size, the difference in average vocal RT is not great—.092 sec. on Session 1 and .062 on Session 3 for all stimuli in each set. The increase in RTs in going from numerals to letters is thus within the range from .054 to .036 sec. per bit, which is greatly below the values reported when less familiar coding alphabets are used.

The generalization appears warranted, therefore, that when *S* is asked to give stereotyped responses to highly familiar sets of symbols, such as those used in English, his information-handling efficiency will already be very high, and his RT will be reduced only slightly by temporary reductions in alphabet size, even when the temporary reduction in uncertainty makes maximum use of familiar subsets.

The present results form the basis for a better understanding of the role of learning in choice behavior. All recent attempts at development of a theoretical model to account for the increase of RT with increases in stimulus uncertainty assume the efficient or selective use of memory (see Welford, 1960), but none deal specifically with the nature of the selectivity of the memory process. This is true both of models which propose that a simultaneous or successive search process (Hick, 1952; Rappaport, 1959) or that a sequential statistical decision process (Stone, 1960; Tanner, 1961) is responsible for the relation between increased time and increased number of alternatives. Thus, Rappaport (1959) assumes that "the task is to match the stimulus perceived with one of the *n* stimuli stored in the memory" (p. 301), and Stone (1961) assumes that *S* is familiar with the probability distributions of the alternative events and that "such familiarity may be the result of a process of learning, provided *S* has performed many trials" (p. 255). The cognitive set theory proposed here is an attempt to specify more explicitly some of the characteristics of the process that makes a selective portion of *S*'s memory available in a particular choice situation. In terms of a computer analogy, the establishment of an appropriate cognitive set for a particular task is like transferring a portion of the total stored information relating to previous signals to a rapid-access memory, and providing appropriate instructions for effecting a given (search or statistical decision) operation which will require comparison of incoming signals with the information stored in the rapid-access memory.

There remains the question of why in certain experiments, particularly those of Leonard (1959), Mowbray and Rhoades (1959), and Mowbray (1960), RT shows no change at all as the size of the set of alternatives is changed. In all three of these experiments responses were either highly familiar and/or compatible ones. It is possible that a different type of choice mechanism becomes operative in such cases. For example, a search or statistical decision mechanism might be replaced by a direct association mechanism that is independent of the number of alternatives. However, it should be noted that in the first two of these studies the same *S*s served under all conditions; thus another possible explanation is that *S*s may have developed an undifferentiated cognitive set across all the experimental conditions, and not changed this cognitive set when *E* changed conditions.

The results from the recent study by Mowbray (1960) are particularly relevant to the present discussion. He trained 150 *S*s, in five groups. Each group served under a single condition (2, 4, 6, 8, or 10 alternatives) for one session and each *S* made a total of 60 responses to each stimulus in his set. Stimuli were numerals and the response was to vocalize the familiar name. All groups, regardless of number of alternatives, gave a mean RT of about .47 sec., and Mowbray (1960) concluded that ". . . the results support the hypothesis that choice reaction times for practiced responses are constant" (p. 200) over the range from 2 to 10 alternatives. The significant fact, however, is that all of the subsets of numerals which Mowbray[2] used were relatively unfamiliar ones: 4, 8 for two-choice; 1, 4, 6, 8, for four-choice; 1, 2, 3, 5, 7, 8, for six-choice; etc. Thus Mowbray's results are actually in full agreement with those obtained in the present experiments and are consistent with the predictions made from a cognitive set theory.

SUMMARY

The investigation concerned *S*'s ability to reduce his reaction time (RT) on the basis of knowledge that only a small subset of a larger, familiar alphabet would occur as stimuli in a particular experiment. It was predicted that *S*s could make effective use of this knowledge only when the smaller subset itself constituted a highly familiar group. English letters and numerals were used as stimuli and familiar as well as unfamiliar subsets were employed; the response was to vocalize the familiar name of the stimulus.

The results of three experiments were in agreement with the predictions. RTs were faster for small, familiar subsets (such as 1, 2; A, B, C); RTs for small, unfamiliar subsets (such as 2, 7; 4, 7; E, B, P) were about as slow as for the larger familiar sets of which they normally were a part. In no instance, however, was the absolute reduction in RT resulting from a decrease in number of alternatives very large for English language alphabets. It therefore appears that large differences in RT may result from changes in number of alternatives only when tasks are low in S-R compatibility, such as tasks using novel S-R codes or requiring novel information transformations.

These results support the view that information-handling rate is in part a function of cognitive sets which reflect the preparation which *S* makes, in advance, for responding to any one of a group of stimuli. Some characteristics of these cognitive sets are discussed, and the notion is related to recent theories of choice behavior.

REFERENCES

BRAINARD, R. W., IRBY, T. S., FITTS, P. M., & ALLUISI, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

CROSSMAN, E. R. F. W. The measurement of discriminability. *Quart. J. exp. Psychol.*, 1955, **7**, 176–195.

DASHIELL, J. F. A neglected fourth dimension to psychological research. *Psych. Rev.*, 1940, **47**, 289–305.

---

[2] G. H. Mowbray, personal communication.

Davis, R., Moray, N., & Treisman, A. Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, 1961, **13**, 78–89.

Fitts, P. M., & Seegar, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

Gibson, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

Helson, H. W. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* New York: McGraw-Hill, 1959.

Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

Howell, W. C., & Kraft, C. L. Size, blur, and contrast as variables affecting the legibility of alpha-numeric symbols. *USAF WADC tech. Rep.*, 1959, No. TR 59–536.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Klemmer, E. T., & Muller, P. F. The rate of handling information. Key pressing responses to light patterns. *USAF HFORL memo. Rep.* 1953, No. 34.

Leonard, J. A. Tactual choice reactions. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

Leonard, J. A. Choice reaction time experiments and information theory. *Proc. 4th Inform. Theory Conf.*, 1961.

Miller, G. A. *Language and communication.* New York: McGraw-Hill, 1951.

Morin, R. E., & Grant, D. A. Learning and performance of a key-pressing task as a function of the degree of spatial stimulus-response correspondence. *J. exp. Psychol.*, 1955, **49**, 39–47.

Mowbray, G. H. Choice reaction time for skilled responses. *Quart. J. exp. Psychol.*, 1960, **12**, 193–202.

Mowbray, G. H., & Rhoades, M. U. On the reduction of choice-reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

Newman, E. B. Men and information: A psychologist's view. *Nuovo Cim.*, 1959, **13**, 539–559.

Postman, L., & Bruner, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369–377.

Pratt, F. *Secret and urgent.* New York: Bobbs-Merrill, 1939.

Pierce, J. R., & Karlin, J. E. Reading rates and the information rate of a human channel. *Bell Sys. tech. J.*, 1957, **36**, 497–516.

Quastler, H., & Brabb, B. Human performance in information transmission. V. The force of habit. *U. Ill. Control Sys. Lab. Rep.*, 1956, No. R-70.

Rappaport, A. A study of disjunctive reaction times. *Behav. Sci.*, 1959, **4**, 299–315.

Seibel, R. Some factors influencing choice reaction time. Paper read at the American Psychological Association meetings, Chicago, 1960.

Stone, N. Models of choice-reaction time. *Psychometrika*, 1960, **25**, 251–260.

Tanner, W. P. Physiological implications of psychophysical data. *Ann. N. Y. Acad. Sci.*, 1961, **89**, 752–765.

Welford, A. T. *Ageing and human skill.* Oxford: Oxford Univer. Press, 1958.

Welford, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, **3**, 189–229.

(Received January 14, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 330–336

APPLICATION OF A STIMULUS SAMPLING MODEL TO CHILDREN'S CONCEPT FORMATION WITH AND WITHOUT OVERT CORRECTION RESPONSES[1]

PATRICK SUPPES AND ROSE GINSBERG

Institute for Mathematical Studies in the Social Sciences, Stanford University

In the present experiment, the first of a series of experiments intended to study concept formation in young children, 5- and 6-yr-old *S*s are required to learn two fairly difficult concepts, each represented by three stimuli. The concepts to be learned are numbers in binary notation and the stimuli representing them are varied by the use of different symbols for the binary digits. A feature of the experimental situation is that it permits the application of a simple sampling model to the data and in addition suggests two methods of analysis, one in terms of paired-associate (PA) learning and the other, which we shall later refer to as "pure property analysis," more directly concerned with acquisition of the concepts involved.

If, in analyzing the present data the six stimuli representing two concepts are treated as if they were independent, the situation is analogous to a PA learning experiment. The data then become amenable to analysis in terms of a statistical learning theory that has been successfully applied to PA experiments (Bower, 1961; Estes, 1961).

In the second analysis the three stimuli representing the same concept are treated as if they were identical—in effect one stimulus. The learning situation is then reduced to the association of two stimuli and the two appropriate responses and the PA model used in the first analysis may again be applied to the rearranged data. In this second method of analysis the learning situation is viewed as involving the association between the common property of the stimuli which determines a concept and the identification of that concept, which is the correct response. To distinguish the two approaches the first method—six independent stimuli and two responses—is designated the "paired-associate" analysis, and the second approach—two independent stimuli and two responses—the "pure property" analysis.

A further aim in the present experiment is to establish the role in young children's learning behavior of an overt correction response following an incorrect response. In two-response situations with adult human *S*s, when *S* is informed whether he is correct or incorrect on each trial, the probability of making a correct response is affected by the outcome of a trial rather than the specific response made on that trial. Requiring *S* to make an overt correction response after reinforcement neither improves his learning rate nor influences his asymptotic behavior (Burke, Estes, & Hellyer, 1954). The present experiment attempted to establish whether an analogous situation obtains when *S*s are young children.

---

[1] The research reported was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare. We are indebted to W. K. Estes for a number of constructive suggestions.

METHOD

Subjects.—There were 48 *S*s divided between two groups. Each group consisted of 12 *S*s from Grade 1 and 12 *S*s from the kindergarten class of Stanford Elementary school. The average age of the former was 6 yr. and 3 mo., of the latter 5 yr. and 1 mo.

Design.—On each trial a stimulus was presented to which *S* made one of two possible responses. The *E* always indicated whether *S*'s choice on that trial was correct or incorrect. For 24 *S*s (Group C) a wrong choice was followed by an overt correction response; for the remaining 24 *S*s (Group NC) no overt correction response was required.

Equipment.—The stimuli, printed symbols on 8½ × 11 in. white cards, were presented one at a time on a small metal reading stand to the front of which a sizeable white hook was attached. Two responses were available, the Arabic numerals 4 and 5 each printed in ¼-in. red tape on a 2½ × 4 in. white card. A 1-in. composition ring attached to the top of each response card enabled *S* to hang the chosen response on the reading stand hook.

Six different stimuli were presented. Three of the stimuli represented 4 in the binary number system, the remaining three represented 5. The equivalent binary numbers are, of course, 100 and 101, but as quite small children are often familiar with some combinations of 1 and 0, different symbols were used in place of the Arabic numerals. The stimuli used and their Arabic numeral equivalents are shown in Table 1. Sixteen copies of each stimulus were prepared, making 96 stimuli in all. Four different random sequences were used and matched across groups.

Procedure.—The *S*s were run individually. Each *S* was seated on a small chair with the reading stand on a low table in front of him. A response card was placed in front of and to the right and left of the reading stand (the position, left or right, of the responses was balanced across *S*s). The *S*s were told that the numbers on the response cards were 4 or 5. All the children were familiar with these Arabic numerals although recognition was not, of course, necessary to complete the task. It was explained that all the "pictures" to be shown were different ways of writing 4 or 5; that is, they were all the "same as" 4 or the "same as" 5. The *S*s were instructed to guess whether a picture presented on the reading stand was the same as 4 or 5 and to indicate their guesses by putting the chosen response on the hook in front of the picture.

TABLE 1

STIMULI AND EQUIVALENT ARABIC NUMERALS USED IN LEARNING AND TEST TRIALS

Trials	Stimuli	Numerals
Learning	Γ** λππ ↔ΣΣ	4
	Γ*Γ λπλ ↔Σ↔	5
Test	δϵϵ αΔΔ	4
	δϵδ αΔα	5

A practice card was then placed on the reading stand; 2 sec. later *E* said "now" and *S* placed the chosen response card on the hook. The *E* said "right" if the answer was correct and "no" if the answer was incorrect. In Group NC (noncorrection) this completed a trial; in Group C (correction), after an incorrect response, the response card was taken from the hook and *S* was given time to place the correct card on the hook. The trial was terminated by *E* saying "right." In both groups at the end of a trial *S* was allowed to remove the response card from the hook while *E* changed stimuli. The correct intertrial position for the response cards was established before the next stimulus was exposed. A fairly constant speed for each movement on each trial was in general easily established both for *S* and *E*. For all *S*s it was not necessary after the first few trials for *E* to initiate *S*'s response by saying "now."

To demonstrate the procedure to *S*, two practice trials were given. Each of the stimuli used on these trials consisted of two different objects (dissimilar also to the symbols used in the learning trials) and the response to the first stimulus was always judged as correct, to the second stimulus as incorrect. The order of presentation of practice stimuli was reversed for half the *S*s across both groups and grades.

The 96 training trials following the practice trials were broken up by a 2-min. rest period between the first and second 48 trials. The learning trials were succeeded by a second 2-min. rest period after which 4 test trials were given. The test stimuli, two of which represented 4 and two represented 5, are shown in Table 1 together with the correct responses. No information of any kind was given on these four trials but at the end of this short session all *S*s were told they had answered correctly on all trials.

RESULTS AND DISCUSSION

Mean learning curves.—As there was no significant difference between the percentages of correct responses for kindergarten and Grade 1 *S*s (the total numbers of correct responses for the two grades being, respectively, 1526 and 1534), the data from both grades were combined and are presented in terms of proportion of errors for Groups C and NC in Fig. 1. The *t* of 4.00 (*df* = 46) computed between overall responses for the two groups is significant at the .001 level. Although Group C does significantly better than Group NC, *S*s in the former group do not achieve perfect learning. As a matter of interest both groups were divided into *S*s who reached a criterion of at least five correct responses on each of the last three presentations of the six stimuli, and *S*s who did not achieve that criterion. In Group C, 6 of the Grade 1 *S*s and 6 of the kindergarten *S*s reached criterion; in Group NC, 2 Grade 1 and 2 kindergarten *S*s reached criterion. Figure 2 describes the correct responses of criterion and noncriterion *S*s in Groups C and NC. The difference in learning rate between *S*s not reaching criterion in Groups C and NC is apparently considerably affected by the correction response variable. A *t* of 2.75 (*df* = 30) between the overall correct responses of these two subgroups was significant at the .01 level. These results, both for the full groups and the subgroups, indicate that the making of an overt correction response has a positive additional reinforcing effect in young children. This is contrary to the Burke et al. (1954) results for adults.

Theory.—Bower (1961) has recently presented a one-element model for PA learning from which he derived an extensive number of predictions. (A mathematically equivalent model was first formulated by Bush and Mosteller, 1959; see also Estes, 1961.) In effect, it is assumed that in a PA experiment each item may be represented by one stimulus element which may be in one of two states, conditioned to the correct response or unconditioned. On each trial the stimulus associated with the

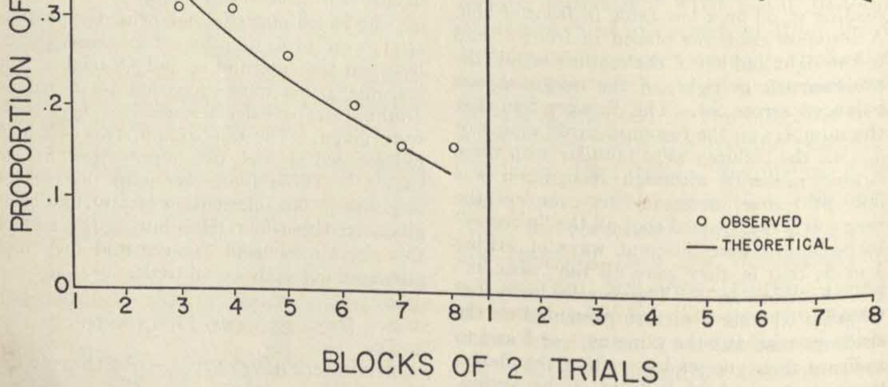


FIG. 1. Observed errors for the 16 trials on each item, averaged over *S*s and items, and predicted curves from the paired-associate analysis.

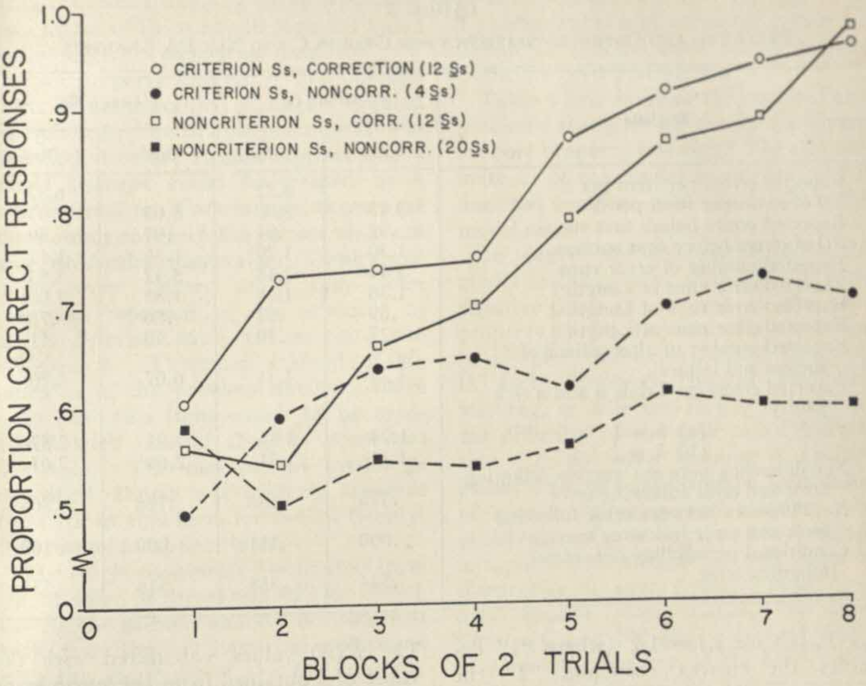


FIG. 2. Proportion of correct responses for *S*s with at least five out of six correct on each of the last three trials (criterion *S*s) and for "noncriterion" *S*s.

item presented is sampled. If the stimulus is unconditioned and the trial is reinforced, it becomes conditioned to the correct response with probability θ and remains unconditioned with probability $1 - \theta$. At the outset of the experiment the stimulus is assumed to be unconditioned. The probability of a response if an unconditioned stimulus is sampled is $1/k$ where k is the number of available responses. The probability of a correct response if a conditioned stimulus is sampled is unity. The reader will recognize that this model is a special case of stimulus sampling theory when the presence of exactly one stimulus element is assumed.

The statistics which Bower derives are presented in some detail and are not therefore developed here. However, since many of the derived expressions assume a larger number of trials or a larger value of θ, they are not appropriate in the present case where both n and θ are small. Consequently wherever necessary we have re-estimated the statistics for finite n and small θ. Some but not all of the formulas we used are given in Estes (1961).

Paired-associate analysis.—Using the one-element model described, with each stimulus treated as an independent item, the single parameter θ was estimated from the number of errors. The value of θ for Group C was .088, for Group NC, .031. The theoretical mean error curves are presented in Fig. 1.

Table 2 presents some predicted and observed statistics for Groups C and NC. For both groups the model, judged from the statistics presented, describes the data rather well. The sums of squared deviations between predicted and observed statistics are nearly the same for both groups, being .118 for Group C and .122 for Group NC.

Certain observed and predicted joint probabilities are given in Table 3. Designating the probability of response i on trial m and response j on trial $m + 1$

TABLE 2

Predicted and Observed Statistics for Groups C and NC: PA Analysis

Statistic	Group C		Group NC	
	Pred.	Obs.	Pred.	Obs.
Expected errors per item per *S*		4.32		6.37
SD of errors per item per *S*	3.13	3.00	3.05	2.97
Expected errors before first success	.92	.94	.97	.89
SD of errors before first success	1.30	1.29	1.36	1.44
Expected number of error runs	2.44	2.48	3.42	3.53
Expected error runs of Length 1	1.36	1.48	1.85	2.00
Expected error runs of Length 2	.59	.52	.86	.78
Expected error runs of Length 3	.27	.20	.40	.41
Expected number of alternations of success and failure	4.25	4.31	6.07	6.20
Expected errors on Trials *n* and *n* + *k* jointly				
(a) *k* = 1	1.94	1.86	2.94	2.87
(b) *k* = 2	1.71	1.51	2.69	2.67
Net difference between success following error and error following error	.375	.486	.188	.306
Net difference between error following error and error following success	.000	.111	.000	.160
Conditional probability of success following error	.544	.558	.516	.525

as $P_m(ij)$ for $i,j = 1,2$ (where "1" denotes the correct response, "2" the incorrect response), and letting:

$$\bar{P}(ij) = \frac{1}{n-1}\sum_{m=1}^{n-1} P_m(ij)$$

it can be shown that:

$$P_m(11) = 1-(1-\theta)^{m-1}(3/4-\theta/4) \qquad [1]$$

and:

$$\begin{aligned}\bar{P}(11) &= 1/(n-1) \qquad [2]\\ &\quad\times \sum_{m=1}^{n-1}[1-(1-\theta)^{m-1}\\ &\qquad\times(3/4-\theta/4)]\\ &= 1-(3/4-\theta/4)\\ &\qquad\times\left[\frac{1-(1-\theta)^{n-1}}{(n-1)\theta}\right]\end{aligned}$$

and similarly:

$$\begin{aligned}\bar{P}(12) = \bar{P}(22) &= (1-\theta) \qquad [3]\\ &\quad\times\left[\frac{1-(1-\theta)^{n-1}}{4(n-1)\theta}\right]\end{aligned}$$

$$\bar{P}(21) = (1+\theta)\left[\frac{1-(1-\theta)^{n-1}}{4(n-1)\theta}\right] \qquad [4]$$

The $\bar{P}(ij)$ values, calculated with the value of θ obtained from the mean error curve, are presented in Table 3 together with the observed $\bar{P}(ij)$. Again the fit for both groups is good. In addition, θ was estimated directly from each of the four observed joint probabilities for Group C. The results were $\hat{\theta}_{11} = .089$, $\hat{\theta}_{12} = .086$, $\hat{\theta}_{21} = .086$, and $\hat{\theta}_{22} = .094$. The mean of these four values is .089, which is surprisingly close to the mean error curve estimate of .088. The closeness of these two estimates is encouraging because in many probabilistic learning experiments the value of θ estimated

TABLE 3

Average Joint Response Probabilities
$P(ij)$

$P(ij)$	Group C		Group NC	
	Predicted	Observed	Predicted	Observed
$\bar{P}(11)$	.587	.589	.399	.394
$\bar{P}(12)$	.129	.131	.196	.202
$\bar{P}(21)$	.155	.156	.209	.212
$\bar{P}(22)$	.129	.124	.196	.192

from the mean learning curve is smaller by a factor of from 3 to 10 than the value estimated from sequential data.

Pure property analysis.—The one-element model described was again applied to the data, but in this instance it was assumed that the stimuli representing a single concept could be treated as a single stimulus. For example, there were three items to which the correct response was the Arabic numeral 4. These three items were treated as if they were identical, representing one stimulus to which *S* must learn to associate the response 4. Therefore, instead of six items as in the previous analysis, there were now two items—each to be associated with one of the two responses available. The one-element model as described above was used to analyze the data in this form for Group C only, for reasons to appear later.

The single parameter θ estimated from the number of errors was .028 for Group C. As the present analysis is concerned with two distinct items, as compared with the six items of the PA analysis, it follows that the method of estimating θ from the mean errors will give a parameter value approximately one-third that of the estimate for the PA approach. The theoretical and empirical curves for Group C, plotted over blocks of four trials, are presented in Fig. 3.

Table 4 lists some of the obtained and predicted statistics for the data analyzed for pure property learning. The approximate fit of the model is not bad, but it does not compare with the fit of the PA model.

In view of the excellent fit of the PA model to the present data and the comparative lack of success of the pure property model, the immediate question is whether the experimental situation, in fact, involved concept formation learning, or whether it was reduced by the difficulty of the task to a PA situation. If *S*s were learning a concept rather than, or in addition to, association of specific stimulus and response, the probability of a correct response when *S* is faced with a completely new stimulus illustrating the same concept should be at least greater than chance. For both Groups C and NC the initial probability of a correct response on the first experimental trial of the original learning session was .5. Over the 4 test trials introduced after the 96 learning trials, the propor-

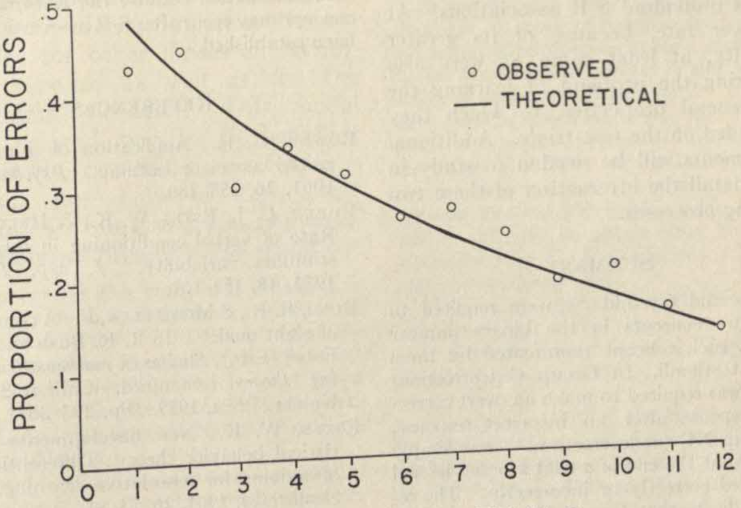


FIG. 3. Observed errors and predicted curve for Group C from pure property analysis.

TABLE 4

Predicted and Observed Statistics for Group C: Pure Property Analysis

Statistic	Predicted	Observed
Expected errors per item per S		13.12
SD of errors per item per S	8.85	6.29
Expected errors before first success	.972	1.06
SD of errors before first success	1.36	1.49
Expected number of error runs	6.78	8.50
Error runs of Length 1	5.76	5.57
Expected number of alternations, successes, and failures	13.06	16.25
Expected errors on Trials n and $n + 1$ jointly	6.34	4.48

tion of correct responses for Group C was .646, for Group NC .479. The obtained value for Group C is significantly different from chance at the .01 level ($t = 2.76$, $df = 23$). The fact that the proportion of correct responses for Group NC on the test trials was at the level of chance is the reason the pure property analysis was not applied to this group.

The better fit of the PA model in comparison with the pure property model for Group C and the somewhat paradoxical evidence from test trials that learning had occurred with respect to the property do not seem inconsistent to us, but rather appear to have a natural psychological explanation. Initially *S*s were mastering the problem by acquiring the six individual S-R associations. At a slower rate, because of its greater difficulty, at least some *S*s were also mastering the problem of learning the two general properties, to which they responded on the test trials. Additional experiments will be needed to study in more detail the interaction of these two learning processes.

SUMMARY

Five- and 6-yr-old *S*s were required to learn two concepts in the binary number system, each concept represented by three different stimuli. In Group C (correction) each *S* was required to make an overt correction response after an incorrect response, in Group NC (noncorrection) *S* was simply informed at the end of a trial whether he had responded correctly or incorrectly. The results indicate that Group C performed significantly better than Group NC.

Two methods of analysis were used. For the paired-associate (PA) analysis the stimuli were treated as if they were independent items and for the pure property analysis, all stimuli describing a single concept were treated as if they were identical items. A simple sampling model for PA learning was applied to the data in each case; a good fit of the model to the data was obtained for the PA analysis, the pure property analysis was performed only with the data of Group C and did not approach the adequacy of the PA analysis.

Test trials on new stimuli following the learning trials provided evidence that some *S*s in Group C were learning the concept involved. It was pointed out that evidence of concept learning is not incompatible with success of a PA analysis in the present experimental situation because the learning of the concept may occur after S-R associations have been established.

REFERENCES

Bower, G. H. Application of a model to paired associate learning. *Psychometrika*, 1961, **26**, 255–280.

Burke, C. J., Estes, W. K., & Hellyer, S. Rate of verbal conditioning in relation to stimulus variability. *J. exp. Psychol.*, 1954, **48**, 153–161.

Bush, R. R., & Mosteller, F. A comparison of eight models. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford, Calif.: Stanford Univer. Press, 1959. Pp. 293–307.

Estes, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika*, 1961, **26**, 73–84.

(Received December 14, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 337–341

MAGNITUDE OF REINFORCEMENT AND CHOICE BEHAVIOR IN CHILDREN[1]

SIDNEY SIEGEL AND JULIA McMICHAEL ANDREWS

Pennsylvania State University

In the classical two-choice uncertain outcome experimental situation, first used by Humphreys (1939), *S* predicts, for a series of trials, which of two events will occur on each trial. The events occur with fixed, unequal probabilities, π and $(1 - \pi)$, in a random sequence, the structure of which is in no way contingent upon *S*'s behavior. By witnessing the events as they occur, *S* sees for himself the accuracy of his predictions.

Both Estes' statistical learning model (Estes & Straughan, 1954) and a decision-making model developed by Siegel (1959) yield predictions of asymptotic choice behavior in this situation. The Estes model predicts that, after learning takes place, *S*s tend to predict the two events in the proportion of their actual occurrence. Siegel's Model I, based on considerations from decision-making theory, provides for other modes of asymptotic response as well as for the "matching" behavior. This model yields the prediction that the behavior of *S*s in the two-choice situation is a function not only of the event probabilities but also of the *amount of reinforcement*[2] present in the situation.

Basic to the Siegel model is the hypothesis of *maximization of expected utility*: a person behaves in a choice situation as if he were attempting to maximize the expected subjective value of the outcome of his behavior. Particularly important, on the basis of this hypothesis, are the two concepts of the utility (subjective value) of a correct response and the utility of variability, or the negative utility of "boredom" (Siegel, 1959, 1961).

The prediction of asymptotic response probability, p, yielded by the model[3] may be expressed as follows (Siegel, 1959, p. 307):

$$p = \alpha(\pi - \tfrac{1}{2}) + \tfrac{1}{2}$$

π = the probability of occurrence of the more frequent event.

p = the proportion of times the subject chooses the more frequent event;

and $$\alpha = \frac{a}{b}$$

a = the marginal utility of a correct prediction.

b = the marginal utility of variability.

As can be seen, the model provides for a change in *S*'s mode of response as an increase in amount of reinforcement for a correct prediction produces an increase in the marginal utility of such a prediction for *S*. It is also seen that in the case when the marginal utility of a correct prediction, a, is equal to the marginal utility of variability, b, the Siegel model predicts the same asymptotic behavior as does the Estes model. Such a case would seem to obtain when knowledge of the outcome of his predictions is *S*'s only source of reinforcement.

Findings in support of both the Estes model and the Siegel model have been reported (e.g., Estes & Straughan, 1954; Siegel, 1961; Siegel & Goldstein, 1959).

---

[1] Support for this study was provided in part by grants to S. Siegel from the National Science Foundation (G-7071) and from the Central Fund for Research of the Pennsylvania State University, and in part by a Public Health Service Research Fellowship to J. M. Andrews from the National Institute of Mental Health.

[2] It will be noted that the term "amount of reinforcement" is used here with reference to the particular reinforcement for a correct response rather than to the probability of reinforcement, π.

[3] A generalization of the Siegel model for k alternatives has been developed (Siegel & McMichael, 1960). The above expression of the prediction of the model may be derived as a special case of the general model, the case for which $k = 2$, with equal amounts of reinforcement for the two alternatives.

The behavior of children in probabilistic situations has been a specific concern of both choice-behavior investigators and developmental psychologists (e.g., Atkinson, Sommer, & Sterman, 1960; Messick & Solley, 1957; Stevenson & Weir, 1959). Among the research orientations of these investigators, as among those of investigators of adult choice behavior, there has been considerable variation in the degree to which amount of reinforcement is considered a relevant or significant variable in the research setting. It is suggested here that a portion of the variability observed both within and between investigations of children's choice behavior may be a function of this variable, amount of reinforcement.

For the present study preschool children have been observed under two conditions of amount of reinforcement in a suitable modification of the two-choice uncertain outcome paradigm in order to determine the applicability of Siegel's model for choice behavior in this situation to the behavior of young children.

It is hypothesized that the behavior is in part a function of the amount of reinforcement provided for a correct response, as well as of the event probabilities. The prediction is that an increase in the amount of reinforcement for a correct prediction, by producing an increase in a child's marginal utility of a correct prediction, leads to an increased tendency for a child to maximize the probability of such a prediction.

METHOD

Subjects.—The *S*s were 24 boys ranging in age from 3–10 to 5–0, a sample including both nursery-school and nonnursery-school children.

Experimental conditions.—The two-choice uncertain outcome situation was presented to each child in the form of a game in which he predicted, for a series of trials, which of two identical containers placed before him contained an object. After each prediction the contents or lack of contents of the chosen container was revealed to the child. The container in one of the positions held the object more frequently than that in the other position.

Two conditions of amount of reinforcement were used. In the low reward condition, the object used in the series was a button, selected to have minimum interest value for the children. The same button was used throughout the series. Knowledge of the outcome of each trial, as verified by *S*'s indication at the end of the trial as to which container had held the object, was the only experimental reinforcement. Use of verbal reinforcement was avoided by *E*.

In the high reward condition, the object was a small prize which *S* could have to keep if he predicted correctly which container held it. The prizes for the entire series included a wide variety of toys, candy, plastic trinkets, crayons, balloons, coins, and other similar items, rewards of the type shown to have high reinforcement value for nursery-school-age children (Bijou & Sturges, 1959). For each trial, then, there was for *S* the possibility of receiving a prize, as well as the knowledge of the outcome of the prediction.

Design.—Each *S* was observed under both conditions of reinforcement. The 24 *S*s were randomly assigned to two groups of equal size, the random assignment determining the first reinforcement condition under which *S* would be observed. Those *S*s observed first under low reward constituted Group L; *S*s observed first under high reward constituted Group H.

There were 100 trials in the two-choice situation in the first condition. The series of events was randomized over blocks of 20 trials with π, the probability of the more frequent event, equal to .75. The same series was used for all *S*s, with half the *S*s in each group being randomly assigned left as the position of the more frequent stimulus and the other half assigned right.

Procedure.—Each *S* was seen individually in his home for one session. In each case *S* had been told that two young women were coming to his home to "play some games" with him. The *S* and *E* sat facing one another at a small table. An assistant sat to one side of *E*.

The *E* began by explaining to *S* that he was to play a guessing game, using the materials placed before him on the table. The basic apparatus was a wooden box, 10 × 14½ in., covered at the end placed nearer *E* by a wooden platform. Attached to the platform were two metal rings, about 4½ in. apart and equidistant from the end of the box, into which were set two identical blue toy milk bottles, about 4½ in. in height.

Each *S* in Group L was then told that *E* would take away the bottles, "hide" a button in one of them, and return them to their positions on the platform, at which time *S* should "guess" where the button was, indicating his prediction by pointing. Instructions for Group H were essentially the same except for the reference to the contents of the bottles. Each *S* in this group was told that *E* would place a "prize" in one of the bottles; it was emphasized to *S* that the prize would be his to keep if he guessed its location correctly. There were two practice trials.

The *E* "hid" the button, under low reward conditions, by putting the bottles behind her back for each trial and dropping the button into one of them, according to the prearranged random series. Under high reward conditions, *E* held on her lap a box of assorted prizes hidden from *S*'s view by the lid of the box, set perpendicular to the bottom of the box and behind which the bottles could be filled. The assistant recorded the responses.

For each trial *S* indicated his prediction by pointing to one of the bottles. Regardless of the correctness of the response, *E* turned the chosen bottle upside down over the part of the box nearer *S*, where the contents, if any, fell. The bottle was then returned to its former position, and *S* was asked "Which one was it in?" to which he responded by pointing.

After the 100 trials in the initial condition, *S*s in Group L were given the same series of 100 trials under high reward conditions. This time they were told that they were to play another game which would be the same as the first except that there would now be a prize in one of the bottles which would be theirs to keep if they guessed its location correctly. The *S*s in Group H were given only 12 additional trials under low reward, since, as was expected, these *S*s were usually reluctant to continue, without reward, an activity for which they had previously received prizes. These *S*s were told that they were to play another game, the same as the first game except that now there would be a button in one of the bottles each time.

Rest periods were provided for both groups after 60 and 100 trials; *S*s in Group L rested, in addition, after 160 trials. Additional "breaks" were provided if requested by *S* or deemed advisable by *E*. A session usually lasted from 1 to 1½ hr.

RESULTS AND DISCUSSION

The means of the response probabilities for all trial blocks for both groups are presented in Table 1. Asymptotic response probability, p_∞, is defined for a 100-trial series as the proportion of predictions of the more frequent event for the last 20 trials; for the 12-trial series under low reward for Group H, the proportion for the complete series is tentatively considered asymptotic response probability.

TABLE 1

MEAN RESPONSE PROPORTIONS FOR 20-TRIAL BLOCKS

Group	Reward Condition	Trials	Trial Blocks				
			I	II	III	IV	V
L	Low	1–100	.58	.67	.59	.65	.75
	High	101–200	.72	.75	.78	.82	.85
H	High	1–100	.52	.76	.72	.78	.83
	Low	101–112	.73				

The hypothesis that preschool *S*s observed under conditions of high reward predict the occurrence of the more frequent of two events a larger proportion of the time than do those observed under low reward conditions, thereby showing a greater tendency to maximize the probability of a correct response, was tested as follows.

First test.—Since all *S*s were observed under both conditions of reinforcement, with counterbalancing of the order of conditions, a difference-score comparison of asymptotic response probabilities under low reward and high reward may be made. As predicted, the mean asymptotic response probability for high reward is higher than that for low reward: $\bar{p}_\infty(\text{H}) = .84$ and $\bar{p}_\infty(\text{L}) = .74$. A Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) shows that the difference is significant ($P < .001$).

It was recognized by the authors that a possible explanation of the higher asymptotes for those *S*s for whom high reward was the second condition of reinforcement is that additional learning was taking place under that condition. However, this seems not to be the case: for those *S*s who had high reward as the first condition, $\bar{p}_\infty$ for high reward was .83; for those *S*s who had high reward as the second condition, $\bar{p}_\infty$ for high reward was .85. A sum of ranks test (Siegel & Tukey, 1960) revealed that this difference is not significant ($P > .49$).

Second test.—Since the difference-score analysis was partially based on the 12-trial response series for the group that had low reward second, an additional comparison of asymptotes for the initial conditions of the two independent groups was made. As predicted, the mean asymptotic response probability was higher for Group H than for Group L at the end of the first 100 trials: $\bar{p}_\infty(\text{H}) = .83$ and $\bar{p}_\infty(\text{L}) = .75$. The differences between groups for the initial condition of reinforcement allow rejection of the null hypothesis at $.10 > P > .05$, as determined by a one-tailed sum of ranks test.[4]

Discussion

The prediction of the Siegel model is concerned specifically with asymptotic response probabilities in the two-choice situation. It is necessary, then, in order for the results of the present study to be interpreted as supporting the model, for *S*s to have exhibited stable state behavior in the experimental sessions. As was mentioned in connection with the first test of the hypothesis, differences between *p*'s for the last 20 trials when high reward occurred first (Group H) and the last 20 trials when high reward occurred second (Group L) are not significant. The fact that response proportions under a particular reinforcement condition for Trials 81–100 are similar to response proportions under the same reinforcement condition for Trials 181–200 is evidence in support of the use of the last 20 trials of a 100-trial series as an index of asymptotic behavior.

It may be concluded, then, that the hypothesis of the Siegel model, that asymptotic response probability is largely a function of amount of reinforcement, is supported by the results of the present study.

Stevenson and Weir (1959) have discussed the hypothesis that an increase in the incentive value of reinforcement leads to a tendency in *S*s to seek a solution yielding consistent reinforcement. In terms of the present study, this hypothesis would lead to the prediction that an increase in amount of reinforcement leads to greater response variability, and thus a smaller proportion of predictions of the more frequent event. Since the opposite occurred, it may be concluded that the present study would not lend support to such an hypothesis.

On the basis of the apparent applicability of the Siegel model to the behavior of young children, it is suggested that an "understanding of probability" greater than that attributed to them by Piaget exists among 4-yr-olds. Piaget (1950) reports that, for the probability task he presents to young *S*s, responses which have maximum probability of being correct seem to occur only rarely among children under 7 yr. of age. The results of the present study, however, hardly seem consistent with such findings. Although the behavior of the children under low reward conditions may be discussed without invoking the notion of a concept of probability, it would seem that the tendency to maximize probability of reinforcement under high reward conditions is, in terms of the criterion used by Piaget (1950), behavior reflecting a tendency to respond to actual "quantitative proportions" of events and hence, on the basis of this criterion, is indicative of at least a rudimentary concept of probability in the children exhibiting it. It is possible that the introduction of rewards into a probabilistic task has led to behavior which may be considered more nearly indicative of children's "understanding of probability" than performance in similar tasks in which no reinforcement is provided. Direct support of such an inference has been reported by Yost, Siegel, and Andrews (1962).

---

[4] The behavior of 4 *S*s, 2 in each experimental group, seemed to indicate an incomplete grasp of the task; asymptotes for these *S*s were .55 or below. With these *S*s omitted in the second test of the hypothesis, differences between groups are more significant ($P < .05$). All *P* values reported in the text are based on all *S*s.

SUMMARY

The classical two-choice uncertain outcome experimental paradigm has been adapted for use with preschool children in a test of the predictions of two models for choice behavior in the situation. Children exhibited a greater tendency to maximize the probability of a correct response as the amount of reinforcement for such a response was increased, thereby lending support to one of the alternative models under consideration. It is concluded that the behavior of children in the two-choice situation is a function of the level of amount of reinforcement for a correct response present in the situation, as well as of the event probabilities.

In addition, it is suggested that the maximizing tendency observed among very young children may be indicative of a more advanced stage in the development of a "concept of probability" in children of this age than that which may be inferred from Piaget's results.

REFERENCES

ATKINSON, R. C., SOMMER, G. R., & STERMAN, M. B. Decision making by children as a function of amount of reinforcement. *Psychol. Rep.*, 1960, **6**, 299–306.

BIJOU, S. W., & STURGES, P. T. Positive reinforcers for experimental studies with children: Consumables and manipulatables. *Child Develpm.*, 1959, **30**, 151–170.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

MESSICK, S. J., & SOLLEY, C. M. Probability learning in children: Some exploratory studies. *J. genet. Psychol.*, 1957, **90**, 23–32.

PIAGET, J. Une experience sur la psychologie du hasard chez l'enfant: Le tirage au sort des couples. *Acta psychol., Amst.*, 1950, **7**, 323–336.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SIEGEL, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

SIEGEL, S. Decision making and learning under varying conditions of reinforcement. *Ann. N. Y. Acad. Sci.*, 1961, **89**, 766–783.

SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

SIEGEL, S., & MCMICHAEL, J. E. Individual choice and strategy behavior: A general model for repeated choices. (Res. Bull. No. 11) University Park, Pa.: Department of Psychology, Pennsylvania State University, December 1960.

SIEGEL, S., & TUKEY, J. W. A nonparametric sum of ranks procedure for relative spread in unpaired samples. *J. Amer. Statist. Ass.*, 1960, **55**, 429–445.

STEVENSON, H. W., & WEIR, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.*, 1959, **57**, 403–412.

YOST, P. A., SIEGEL, A. E., & ANDREWS, J. M. Nonverbal probability judgments by young children. *Child Develpm.*, 1962, in press.

(Received December 16, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 342–352

STUDIES IN ASSOCIATIVE INTERFERENCE[1]

NORMA F. BESCH, VENAN E. THOMPSON, AND ALLAN B. WETZEL
Columbus Psychiatric Institute and Hospital, Ohio State University

Twedt and Underwood (1959) have made a distinction between the experimental designs used in the study of verbal proactive transfer in terms of the types of lists used. In the design using "mixed lists," different relationships hold between subsets of items in Lists 1 and 2. Thus, with one group of *S*s learning only two lists, two or more transfer paradigms are being investigated simultaneously. In "unmixed lists," the same relationship holds between all items in Lists 1 and 2 or one paradigm is represented in the two lists. Thus, in a design using unmixed lists, in which two or more proactive paradigms are being compared, each group learns a different set of two lists.

Using mixed lists, Besch and Reynolds (1958) found negative transfer in two paradigms when they were compared to an A-B, D-C control paradigm (the stimuli and responses are both different). The negative transfer paradigms were: (*a*) A-B, A-C (the stimuli are the same but the responses are different), and (*b*) A-B, A-Br (the stimuli and responses are the same but are paired differently). It was suggested that the increase in the negative transfer found for the A-B, A-C paradigm over that usually found with unmixed lists was not due to the greater precision gained with the use of mixed lists, but to the mixing itself. Generalization of habit strength was the means by which the increase was to have occurred. In brief, by the occurrence of stimulus generalization within the first list, associative tendencies were established between the stimuli of the A-B, A-C paradigm and the responses of the A-B, A-Br paradigm. These tendencies, which were further strengthened by generalization during transfer task learning, competed effectively with the correct tendencies, the A-C pairings, on the transfer task.

For the A-B, A-Br paradigm, it was assumed that the associative tendencies established in List 1 were being further strengthened by generalization during the transfer task (List 2). These associative tendencies competed effectively during transfer task learning with the correct tendencies, the A-Br pairings. However, the A-Br pairings, unlike the A-C pairings, were of some strength at the beginning of List-2 learning due to the effects of generalization during List-1 learning. By this interpretation, negative transfer for the A-B, A-Br paradigm would not be a function of list type, since the relationships believed to be critical are the same in both mixed and unmixed lists.

The purpose in the present experiment was to re-examine the generalization interpretation of associative interference by studying pair types intermediate to the A-Br and A-C transfer pair types. Two features distinguish the A-Br from the A-C pair: (*a*) the stimulus of an A-Br pair is paired with a List-1 response which was incorrect for that stimulus on List 1, and (*b*) the response, which was paired with the stimulus of the A-Br pair in List 1, reappears in the transfer task and is paired with another stimulus from List 1. Thus, one intermediate pair type will be formed by pairing a List-1 stimulus with the response term of another List-1 pair. The other type

---

[1] This research was supported in part by Grant G-9061 from the National Science Foundation.

will be formed by pairing a new response with a stimulus from a List-1 pair; the response term of that List-1 pair will be paired with a stimulus from a different List-1 pair. The latter paradigm will be represented as A_1-B_1, A_1-C and the former as A_2-B_2, A_2-B_1.

Using mixed lists, the A_1-C, A_2-B_1, A-Br, and D-C transfer pair types will be compared in Exp. 1, and the A_1-C, A_2-B_1, A-C, and D-C transfer pair types will be compared in Exp. 2 and 3.

EXPERIMENT 1

Method

Subjects.—The *S*s were 40 female volunteers from the introductory psychology course at Ohio State University. They were assigned in equal numbers of one of four counterbalancing groups. An attempt was made in the assignment of *S*s to obtain approximately equal mean scores for the different groups on a practice list.

Design.—The *S*s in each group learned in order a practice list, a pretraining list, and two experimental lists, Lists A and B. Lists A and B were the lists representing the proactive paradigm. The pretraining list was used in an attempt to minimize the effects of word familiarity on the learning of List B, the transfer task. The pretrained words were those words which were present in List B and not present in List A. Since there were four forms of List A, there were also four forms of the pretraining list.

List B, the transfer task, which was identical for all groups, is best considered as consisting of four sets of three word pairs each. The four forms of List A, each learned by a different group of *S*s, were constructed to allow each of these four sets of three pairs to serve in turn under each of the transfer task conditions. The relationships between the word pairs of the four forms of List A and those of List B are presented in Table 1. The relationship between any pair set within Form A and the corresponding transfer pair set, which appears in the last column of the same row, is represented in each case by the paradigm name. The four paradigms are: (*a*) A-B, D-C (three of the pairs of List A were replaced by new S and R terms), (*b*) A-B, A-Br (the S and R terms from three of the List-A pairs were repaired), (*c*) A_2-B_2, A_2-B_1 (three of the stimuli from List A were paired with the R terms from three other List-A pairs), and (*d*) A_1-B_1, A_1-C (new R words were paired in the transfer task with stimuli from three other List-A pairs, the A_2-B_2 pairs). To see clearly the relationship between the A_1-B_1, A_1-C, and the A_2-B_2, A_2-B_1 paradigms, it is necessary to examine in Table 1 more than the sets within the same row. For any A_2-B_2 set in List A, the response words of the corresponding transfer pair set are the response words of the A_1-B_1 set in List A.

Lists and apparatus.—The practice list contained 15 paired one-syllable nouns of low associative strength. List B and the four forms of List A each consisted of 12 pairs of two-syllable adjectives from the Haagen (1949) list. The word pairs for the transfer task were chosen in a way which, in the writer's opinion, minimized meaningful associations, synonymities, and formal similarities both within and between pairs. Thus, List A pairs having words in common with the transfer task met these same conditions. As shown in Table 1, the List A word pairs serving under the A-B, D-C condition and the List A response words serving under the A_2-B_2, A_2-B_1 condition were identical in the four forms of List A. These words were again chosen to satisfy the requirements for the transfer task.

Each counterbalancing group, which learned a different form of List A, also learned a different form of the pretraining list. The response words of each pretraining list consisted of all words which were present in List B and not present in the appropriate form of List A. The stimulus terms of each pretraining list were the first two letters of their paired response words.

A Hunter Model 340 Card Master was used to present all tasks. The 6-sec. interstimulus interval included a 2-sec. stimulus exposure, a 2-sec. stimulus and response exposure, and a 2-sec. blank. A 6-sec. rest interval was given between successive presentations of a list. Three presentation orders were used for each list.

Procedure.—Each *S* learned by the anticipation method four lists in the following order: (*a*) a practice list, (*b*) a pretraining list, (*c*) List A, and (*d*) List B. Six trials each were given on both the practice list and the pretraining list. All *S*s received 12 trials on List A. A criterion of 2 consecutive errorless trials was set for the transfer task; however, if *S* had not reached criterion by Trial 15, practice was discontinued. In the analyses of the data, the assumption was made that if *S*,

TABLE 1

Design for Experiment 1

List A				List B
Form 1 (Group 1)	Form 2 (Group 2)	Form 3 (Group 3)	Form 4 (Group 4)	(All Groups)
A-B, D-C	A-B, A-Br	A_2-B_2, A_2-B_1	A_1-B_1, A_1-C	
S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_1—R_2 S_2—R_3 S_3—R_1	S_1—R_{16} S_2—R_{17} S_3—R_{18}	S_1—R_4 S_2—R_5 S_3—R_6	S_1—R_1 S_2—R_2 S_3—R_3
A-B, A-Br	A-B, D-C	A_1-B_1, A_1-C	A_2-B_2, A_2-B_1	
S_4—R_5 S_5—R_6 S_6—R_4	S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_4—R_1 S_5—R_2 S_6—R_3	S_4—R_{16} S_5—R_{17} S_6—R_{18}	S_4—R_4 S_5—R_5 S_6—R_6
A_2-B_2, A_2-B_1	A_1-B_1, A_1-C	A-B, D-C	A-B, A-Br	
S_7—R_{16} S_8—R_{17} S_9—R_{18}	S_7—R_{10} S_8—R_{11} S_9—R_{12}	S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_7—R_8 S_8—R_9 S_9—R_7	S_7—R_7 S_8—R_8 S_9—R_9
A_1-B_1, A_1-C	A_2-B_2, A_2-B_1	A-B, A-Br	A-B, D-C	
S_{10}—R_7 S_{11}—R_8 S_{12}—R_9	S_{10}—R_{16} S_{11}—R_{17} S_{12}—R_{18}	S_{10}—R_{11} S_{11}—R_{12} S_{12}—R_{10}	S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_{10}—R_{10} S_{11}—R_{11} S_{12}—R_{12}

who had reached criterion, had continued practice, his performance would have been errorless for the remainder of the 15 trials. Ten of the 40 *S*s reached criterion before Trial 15; all of these *S*s reached the criterion between Trials 10 and 15.

Standard instructions for paired-associate learning were given before the presentation of the practice list. The *S*s were informed before each succeeding task that the same procedure was to be followed in learning the new list. The *S*s were not informed concerning the relationship of the pretraining list to the other lists. The time interval between tasks was approximately 3 min.

Results

A simple analysis of variance of the number of correct responses on List A yielded an *F* of less than 1.00 for the main effect of lists. In addition, an analysis of variance of correct responses for the List-A pairs which were identical in the four forms, that is, the pairs serving under the A-B, D-C condition, gave the same result.

Figure 1 shows the percentages of correct responses on each transfer trial for each pair type. The analysis of variance (Lindquist, 1953, Type VI) of the number of correct responses for all groups over the 14 anticipation trials is summarized in Table 2. Due to the significant interactions between

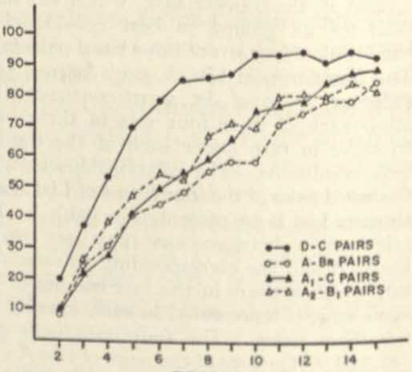


FIG. 1. Percentage of correct responses in Exp. 1 for pair conditions in Trials 2–15 of transfer task.

TABLE 2

Analysis of Variance of Correct Responses for Pair Conditions on Trials 2–15 of the Transfer Task: Exp. 1

Source	*df*	*MS*	*F*
Total between *S*s	39	11.68	
Groups (G)	3	2.60	.21
Error (between)	36	12.44	
Total within *S*s	2200	1.04	
Pair conditions (P)	3	51.14	23.57***
Trials (T)	13	81.10	193.10***
P × T	39	.84	2.33***
P × G	9	5.91	2.72**
T × G	39	.44	1.05
P × T × G	117	.45	1.25
Residual	1980	.47	
Error_1	108	2.17	
Error_2	468	.42	
Error_3	1404	.36	

** $P < .01$.
*** $P < .001$.

Pairs and Trials and between Pairs and Groups, *t* tests of the pair differences were performed for Groups and Trials separately. Performance under Cond. D-C was significantly superior ($P < .05$) to Cond. A-Br on Trials 4–11, to Cond. A_1-C on Trials 4–9, and to Cond. A_2-B_1 on Trials 5, 6, 7, and 10; no other differences were significant. The means and *SD*s of the performance for each pair condition for the four counterbalancing groups are shown in Table 3. The means for Cond. D-C were significantly larger ($P < .05$) than those for all other pairs in all but three comparisons, which were D-C vs. A_1-C and D-C vs. A_2-B_1 for Group 3 and D-C vs. A-Br for Group 2. In addition, for Group 2 the A_2-B_1 mean was significantly smaller than the A-Br mean, and for Group 3 the A_2-B_1 mean was significantly larger than the A-Br mean. Thus, it may be concluded tentatively that in comparison to the D-C pairs, all three of the other pair conditions show negative transfer, at least for the middle portion of learning.

Finally, the intruding responses on the transfer task for each pair condition are classified by source in Table 4. Only a few stimulus words intruded; these were placed in the "Other" category. The number of intruding responses for the pair conditions appears to follow inversely the order of correct responses on the pairs. However, it should be noted that while the difference between the number of intrusions for the A_2-B_1 and A_1-C transfer pairs was large, the difference for correct responses was small.

EXPERIMENT 2

Method

Subjects and design.—Forty female *S*s were selected, assigned to groups, and given practice on four lists in the same manner as in Exp. 1. In Exp. 2, the A-B, A-C paradigm replaced the A-B, A-Br paradigm of Exp. 1. This was accomplished by selecting three new response words to replace the response words which were used in Exp. 1 in the pair sets of the four forms of List A which served under Cond. A-B, A-Br.

Lists, apparatus, and procedure.—The practice list and the transfer list of Exp. 2

TABLE 3

Means and *SD*s of Total Correct Responses on Transfer Task in Exp. 1

Groups	Pair Condition							
	D-C		A-Br		A_1-C		A_2-B_1	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	35.2	4.4	23.7	8.8	26.5	7.4	25.5	8.1
2	32.3	6.7	27.5	7.0	24.5	8.6	22.6	9.3
3	28.9	5.6	19.6	8.0	24.4	5.0	29.2	10.5
4	34.1	6.0	20.9	10.2	24.4	7.6	24.7	7.2

TABLE 4

SOURCE AND NUMBER OF INTRUDING RESPONSES ON THE TRANSFER TASK OF EXP. 1 FOR EACH PAIR CONDITION

Source	Transfer Pair Condition			
	D-C	A-Br	A_1-C	A_2-B_1
A-B, D-C				
List B	26	13	12	13
A-B, A-Br				
Correct (List A)	—	109	—	—
Incorrect (List A)	10	8	13	16
A_1-B_1, A_1-C				
Correct (List A)	—	—	90	—
Incorrect (List A)	—	—	22	—
List B	15	9	9	23
A_2-B_2, A_2-B_1				
Correct (List A)	—	—	—	7
Incorrect (List A)	1	1	4	0
List B	10	15	—	28
Other	12	13	9	9
Unidentified	1	8	10	10
Total	75	176	169	106
Mean	1.9	4.4	4.2	2.6
SD	2.2	3.2	2.4	2.4

Note.—$N = 40$.

were identical to those of Exp. 1. With the exception of one pair set in each form, the four forms of List A were the same in both experiments. In Exp. 2, the same three response words were used in all forms of List A to replace the response words of the pair sets of Cond. A-B, A-Br in Exp. 1. Again, these response words were chosen to minimize, in the four forms of List A, meaningful associations, synonymities, and formal similarities both within and between pairs.

Since all words not present in List A and present in List B were used to construct the pretraining lists, three new pairs were added to the pretraining list of Exp. 1 to construct the lists for Exp. 2.

The apparatus and the procedures used in Exp. 1 were unchanged in Exp. 2.

Twenty-one of the 40 *S*s reached criterion on the transfer task before Trial 15; of these *S*s, 6 reached criterion before Trial 10 and the remainder after Trial 10.

Results

Simple analyses of variance of List-A correct responses and of correct responses on the List-A pairs of the A-B, D-C pair sequence yielded *F*s of 1.52 ($P > .20$) and 1.18 ($P > .20$), respectively, for the main effect of Lists.

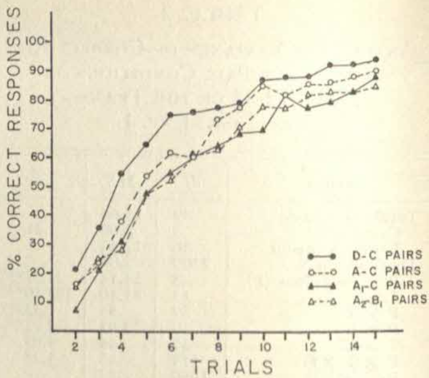


FIG. 2. Percentage of correct responses in Exp. 2 for pair conditions in Trials 2–15 of transfer task.

Figure 2 shows the percentages of correct responses on each transfer trial for each pair type. The analysis of variance of the number of correct responses for all groups over the 14 anticipation trials is summarized in Table 5. Due to the significant interaction between Pairs and Groups, *t* tests were performed for the pair differences within each group. The means and *SD*s of the correct responses for the transfer pair conditions for each of the four groups are pre-

TABLE 5

ANALYSIS OF VARIANCE OF CORRECT RESPONSES FOR PAIR CONDITIONS ON TRIALS 2–15 OF THE TRANSFER TASK: EXP. 2

Source	*df*	*MS*	*F*
Total between *S*s	39	18.25	
Groups (G)	3	26.95	1.54
Error (between)	36	17.52	
Total within *S*s	2200	.91	
Pair conditions (P)	3	18.20	18.76***
Trials (T)	13	79.54	162.33***
P × T	39	.39	1.18
P × G	9	3.54	3.65***
T × G	39	.80	1.63*
P × T × G	117	.37	1.12
Residual	1980	.40	
Error_1	108	.97	
Error_2	468	.49	
Error_3	1404	.33	

* $P < .05$.
*** $P < .001$.

TABLE 6

MEANS AND *SD*s OF TOTAL CORRECT RESPONSES ON TRANSFER TASK IN EXP. 2

Groups	Pair Condition							
	D-C		A-C		A_1-C		A_2-B_1	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	33.3	7.8	25.2	8.0	22.2	10.3	22.5	4.6
2	32.2	7.9	33.9	4.2	28.5	10.2	28.8	5.5
3	26.5	7.6	23.4	6.7	21.8	9.6	24.2	6.0
4	31.1	12.1	28.6	7.9	28.6	9.1	27.1	6.2

sented in Table 6. In comparison to the D-C pairs, the A_1-C and the A_2-B_1 pairs show negative transfer in three out of four of the comparisons. The mean differences were not significant for D-C vs. A_2-B_1 and D-C vs. A_1-C for Groups 3 and 4, respectively. Only in Group 1 was a significant *t* obtained for A-C vs. D-C. Only one A-C vs. A_1-C and one A-C vs. A_2-B_1 comparison were significant; both in Group 2. The significant Trials × Groups interaction indicates that the groups approach their asymptotes at different rates. Since there is little interest in this finding, *t* tests of the group differences over trials were not performed.

TABLE 7

SOURCE AND NUMBER OF INTRUDING RESPONSES ON THE TRANSFER TASK OF EXP. 2 FOR EACH PAIR CONDITION

Source	Transfer Pair Condition			
	D-C	A-C	A_1-C	A_2-B_1
A-B, D-C				
List B	15	21	7	12
A-B, A-C				
Correct (List A)	—	5	—	—
Incorrect (List A)	0	1	0	1
List B	33	4	8	23
A_1-B_1, A_1-C				
Correct (List A)	—	—	91	—
Incorrect (List A)	—	—	27	—
List B	19	23	11	36
A_2-B_2, A_2-B_1				
Correct (List A)	—	—	—	4
Incorrect (List A)	3	2	1	1
List B	1	22	—	13
Other	17	9	21	5
Unidentified	0	8	6	5
Total	88	95	172	100
Mean	2.2	2.4	4.3	2.5
SD	2.5	2.4	3.9	2.6

The intruding responses on the transfer task for each pair condition are classified by source in Table 7. Again, intruding stimulus words are placed in the "Other" category. As in Exp. 1, the number of intruding responses is inversely related to the number of correct responses for each pair condition. Again the differences were not of comparable magnitudes for intrusions and correct responses.

EXPERIMENT 3

The pretraining procedures used in Exp. 1 and 2 may not have affected the different pair types in the same way. Therefore, in Exp. 3, Exp. 2 was replicated over two conditions of pretraining, one relevant and the other irrelevant.

Method

Subjects.—The *S*s were 48 women volunteers from the introductory psychology course at Ohio State University. They were assigned in equal numbers to one of eight groups. Again, the assignment of *S*s to groups was made on the basis of practice list scores.

Design and lists.—The four groups, which served in the relevant pretraining condition, were trained in the same way and learned the same lists as *S*s in Exp. 2. The four irrelevant pretraining groups differed in one aspect from the other groups; the words of the pretraining lists did not appear in any of the subsequent experimental lists. For the four forms of the irrelevant pretraining lists, words were used with the same beginning letters as the words used for the comparable relevant pretraining lists. Where there were repetitions in the words used in the different relevant pretraining lists, repetitions were likewise present across the lists for the irrelevant pretraining groups. Again, the words were selected from the Haagen (1949) list to minimize meaningful associations and synonymities both within the pretraining lists and between it and the other lists.

Apparatus and procedure.—The Lafayette memory drum, Model 303A, which was used to present the practice and transfer lists, operated by exposing with each turn a stimulus word or a stimulus and response word together. The Lafayette Model 303B, which was used to present the remainder of the lists, had an automatic shutter which alternately covered and uncovered the response word of a pair. In the use of both drums, the interstimulus interval was 4 sec. as opposed to the 6-sec. interval of Exp. 1 and 2. The Lafayette drums do not include a blank between pairs as does the Hunter Card Master.

All other procedures used in Exp. 2 were unchanged in Exp. 3.

Of the 11 *S*s in the relevant pretraining groups who reached criterion earlier than Trial 15, 4 reached it before and 7 after Trial 10. For the 7 *S*s of the irrelevant pretraining groups reaching an early criterion, 2 reached it before and 5 after Trial 10.

Results

A factorial analysis of variance of the number of correct responses on List A yielded nonsignificant *F*s for the main effect of Lists, the main effect of Pretraining Groups, and the Groups × Lists interaction. The factorial analysis of correct responses on the List-A pairs serving under Cond. A-B, D-C also yielded nonsignificant *F*s.

Figures 3 and 4 present, respectively, for the relevant and the irrelevant pretraining groups the percentages of correct responses on each transfer trial for each pair type. Table 8 presents the means and *SD*s of the correct responses for the transfer pair conditions for each of the counterbalancing groups of both the irrelevant and the relevant pretraining conditions.

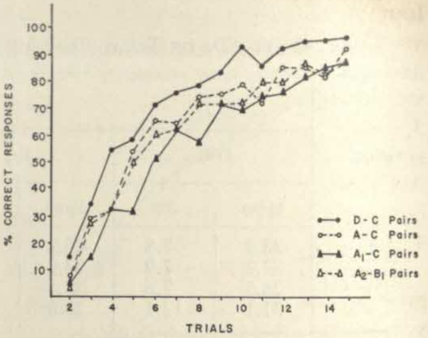


FIG. 3. Percentage of correct responses for pair conditions in Trials 2–15 of transfer task for relevant pretraining groups of Exp. 3.

For the relevant pretraining groups, analysis of variance (Lindquist, 1953, Type VI) yielded results similar to those found in Exp. 2; the main effects of Pairs and Trials were significant ($P < .001$), but only the Pairs × Groups interaction was significant ($P < .05$). The *t* tests, warranted by the significant interaction, gave results different from those of Exp. 2 in only two comparisons;

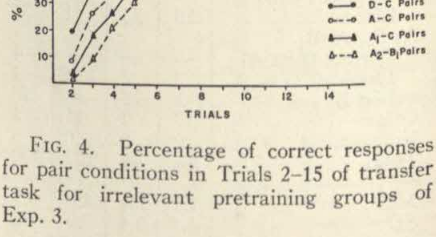


FIG. 4. Percentage of correct responses for pair conditions in Trials 2–15 of transfer task for irrelevant pretraining groups of Exp. 3.

four rather than three of the A_1-C vs. D-C comparisons were significant, and one rather than none of the A_1-C vs. A_2-B_1 comparisons, that for Group 3, was significant. However, the single significant *t*'s for the D-C vs. A-C, A-C vs. A_1-C, and A-C vs. A_2-B_1 comparisons were obtained for Groups 4, 3, and 1, respectively.

For the irrelevant pretraining groups, the same type of analysis of variance yielded significant main effects only, that for Pairs ($P < .001$) and Trials ($P < .001$). The *t* tests for pairs, ignoring counterbalancing groups, indicated that all three experimental pairs differed significantly from the control (D-C) pairs.

From an inspection of Fig. 3 and 4, it appears that the type of pretraining does affect pair performance differently on the transfer task. Thus, the two pretraining conditions were compared with each other in an analysis of variance which ignored the trials factor. An analysis of variance (Lindquist, 1953, Type III) of the total number of correct responses for each pair type for each group was performed. Even though the relevant pretraining groups showed better performance on the experimental pairs and poorer performance on the D-C control pairs in comparison to the irrelevant pretraining groups, the Pairs × Pretraining Groups interaction was not significant ($F = 2.08$, $P < .20$). Only the main effect of Pairs and the Pairs × Counterbalancing Groups interactions were significant ($P = .001$ and .05, respectively).

The intruding responses by pair source for the relevant and irrelevant pretraining groups are presented in Table 9. The intruding responses did not relate inversely to the correct responses as in Exp. 2. The same order for correct responses held for the relevant pretraining groups, but the order of intrusions was changed by the A_2-B_1 pairs which had very few intrusions, in fact, fewer than the D-C pairs. Again, performance was poorest for the A_1-C pairs, along with the largest number of intrusions. In the irrelevant pretraining conditions,

TABLE 8

MEANS AND *SD*s OF TOTAL CORRECT RESPONSES ON TRANSFER TASK IN EXP. 3

Groups	Pair Condition							
	D-C		A-C		A_1-C		A_2-B_1	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Relevant Pretraining								
1	31.8	3.7	29.5	3.6	25.5	7.0	24.2	7.1
2	27.7	12.7	22.8	14.4	22.0	10.7	22.2	10.8
3	31.3	5.6	32.0	5.3	23.2	9.3	32.2	5.7
4	32.5	6.5	23.7	8.3	25.3	9.5	25.8	7.4
Irrelevant Pretraining								
1	32.7	7.2	24.8	5.9	23.0	6.1	17.8	7.8
2	34.2	5.3	26.0	9.5	23.3	7.6	22.3	7.5
3	33.2	5.8	27.0	8.5	22.0	9.3	28.7	6.9
4	28.8	4.2	25.5	9.5	29.5	4.7	21.7	7.6

TABLE 9

Source and Number of Intruding Responses on the Transfer Task for Each Pair Condition for the Relevant and Irrelevant Pretraining Groups

Source	Transfer Pair Condition							
	D-C		A-C		A_1-C		A_2-B_1	
	R[a]	I[b]	R[a]	I[b]	R[a]	I[b]	R[a]	I[b]
A-B, D-C								
List B	7	15	11	2	5	0	5	3
A-B, A-C								
Correct (List A)	—	—	1	1	—	—	—	—
Incorrect (List A)	0	0	0	0	1	0	0	0
List B	16	6	9	9	5	10	8	0
A_1-B_1, A_1-C								
Correct (List A)	—	—	—	—	55	59	—	—
Incorrect (List A)	—	—	—	—	12	6	—	—
List B	5	3	14	12	10	5	8	13
A_2-B_2, A_2-B_1								
Correct (List A)	—	—	—	—	—	—	3	6
Incorrect (List A)	0	0	0	1	0	0	1	0
List B	7	2	8	8	—	—	6	12
Other	6	1	5	2	4	1	2	1
Unidentified	3	1	4	2	6	4	4	3
Total	44	28	52	37	98	85	37	38
Mean	1.8	1.2	2.2	1.5	4.1	3.5	1.5	1.6
SD	2.9	2.0	3.1	2.1	4.3	3.3	2.2	1.6

Note.—$N = 24$.

[a] R refers to relevant pretraining groups.

[b] I refers to irrelevant pretraining groups.

for the first time performance was poorer for the A_2-B_1 pairs in comparison to the A_1-C pairs, but they did not surpass the A_1-C pairs in the number of intruding responses. The number of intrusions was lower for each pair type, with the exception of the A_2-B_1 pairs, for the irrelevant in comparison to the relevant pretraining condition. One would expect the A_2-B_1 pairs to be unaffected since they were not involved in the pretraining.

Discussion

In agreement with previous findings (Besch & Reynolds, 1958; Porter & Duncan, 1953; Twedt & Underwood, 1959), negative transfer was shown for A-Br pairs. For the irrelevant pretraining groups in Exp. 3, the A-C pairs were significantly different from the D-C pairs, while in Exp. 2 and for the relevant pretraining groups in Exp. 3, the A-C pairs did not differ significantly from the D-C pairs. In all of the comparisons between the A-C and the D-C pairs, the differences were in the direction shown in comparable experiments (Besch & Reynolds, 1958; Spiker & Holton, 1958; Twedt & Underwood, 1959). Since Underwood, Runquist, and Schulz (1959) showed that pretraining of the response words of a paired-associate list facilitated subsequent performance on that list, it seems probable that the A-C pairs in the present experiments were facilitated by relevant pretraining. While the difference obtained in Exp. 3 was in the appropriate direction, it was not statistically significant. Therefore, one can conclude only tentatively that the pretraining of both stimuli and responses, as in the D-C pairs, either depresses performance on the subsequent pairings, or facilitates performance to a minor degree in comparison to the facilitation obtained with response pretraining. Finally, both the A_1-C and A_2-B_1 pairs showed negative transfer in all experiments, with the performance levels for these pairs above the level for the A-Br pairs and below that for the A-C pairs, but not significantly so in either case.

The primary purpose underlying the present experiments was to determine whether generalization is a sufficient principle in accounting for negative transfer effects. In an earlier paper (Besch & Reynolds, 1958), in which a "response competition" interpretation was offered, it was implicitly assumed that as long as the response items from the first list were not present in the transfer task (A-B, A-C paradigm), their effects were negligible and could be ignored. However, Spiker and Holton (1958) demonstrated that such is not the case; in the A-B, A-C paradigm, negative transfer, or at least a large portion of it, is due to competition from first task responses. In addition, it also seems clear from the results of Barnes and Underwood (1959) that the A-B stimulus and response associations are gradually extinguished during the course of A-C learning. As Postman (1959) suggests, any account of interference in verbal learning must encompass "unlearning of first task responses" as well as "response competition." Earlier, Melton and Irwin (1940) made this same suggestion. With these considerations in mind, an attempt will be made to evaluate the generalization interpretation of interference by comparing the A_1-C and the A_2-B_1 pairs with the A-C and the A-Br pairs.

According to a generalization interpretation, what would be expected for an A_1-C pair in comparison to an A-Br pair? The previously correct responses on List A for both pair types are present in the transfer task. By the process of generalization of habit strength, these tendencies would be gradually strengthened throughout transfer task learning and thus be maintained at a sufficient strength to compete successfully with the correct tendencies. These interfering tendencies would not be competing with correct tendencies of equal strength, however. The correct tendency for the A-Br pair would be of some strength at the beginning of transfer task learning due to the generalization of associative strength on the first task (List A), while the correct tendency for the A_1-C pair would be of zero or negligible associative strength. The A_1-C pairs, it will be recalled, were easier rather than more difficult than the A-Br pairs.

For the A_2-B_1 pair as for the A-Br pair, the correct tendency on the transfer task has some strength gained from generalization during first-task learning. The previously correct tendency is being strengthened for the A-Br pair, while it is being extinguished for the A_2-B_1 pair. The A_2-B_1 pairs, as would be predicted, were easier than the A-Br pairs.

The A_1-C and A-C pairs would not differ from each other in the strength of the correct tendencies on the transfer task. The A_1-C pairs were more difficult than the A-C pairs, as would be predicted, since the previously correct tendency is being extinguished for the A-C pair, and being strengthened for the A_1-C pair during transfer task learning.

The A_2-B_1 pair differs in one respect from the A-C pair; its correct tendency is of above zero strength due to generalization during first task learning, while that for the A-C pair is of zero or negligible strength. However, the A_2-B_1 pairs were more difficult, rather than easier, than the A-C pairs.

In summary, the direction of the differences shown in half of the comparisons was opposite to that predicted. While these differences were not statistically significant, it does seem reasonable to place confidence in them, particularly since the findings of Exp. 2 were replicated in Exp. 3. It may be concluded that the generalization interpretation, offered in the present paper, by itself does not adequately explain the findings.

It appears necessary to assume that at some point during the learning of any list, associations between nonpaired stimulus and response terms undergo extinction or inhibition. It may be recalled that Gibson (1940) found a similar assumption useful. She hypothesized that, in addition to generalization, a simultaneous process of differentiation is occurring which decreases generalization. Although Gibson did not state how generalization was decreased, she did state that differentiation in verbal learning is analogous to differential inhibition in classical conditioning. An extinction hypothesis, used alone or as used by Gibson, is consistent with the results found in the present experiments. For example, in the A_2-B_1 pair, the correct associations on Task B have been subject to extinction effects on Task A, and thus the probability is decreased that a stimulus of an A_2-B_1 pair will elicit its paired response. This is also true of the A-Br pair. For the A-Br pair, as well as the A_1-C pair, its previously correct response is present in the list; the tendency has some strength and therefore needs to be extinguished during transfer task learning. The situation in the transfer list for the A-Br pairs and the A_1-C pairs would thus be analogous to the existence of preexperimental associations in the Task A. In conclusion, it is felt that an extinction interpretation of interference has some merit, since it would predict the A-C pairs to show the least interference, the A-Br pairs to show the most interference, and the A_1-C and A_2-B_1 pairs to be intermediate to the A-C and A-Br pairs.

SUMMARY

Three experiments were performed for the purpose of evaluating a generalization interpretation of interference. In Exp. 1, four transfer paradigms were compared: *(a)* A-B, D-C, *(b)* A-B, A-Br, *(c)* A_1-B_1, A_1-C, and *(d)* A_2-B_2, A_2-B_1. In comparison to D-C transfer pair learning, all other pairs showed negative transfer. No other differences were significant.

In Exp. 2, four transfer paradigms were compared: *(a)* A-B, D-C, *(b)* A-B, A-C, *(c)* A_1-B_1, A_1-C, and *(d)* A_2-B_2, A_2-B_1. Only the A_1-C and A_2-B_1 pairs showed negative transfer.

In Exp. 3, Exp. 2 was replicated over two conditions of pretraining, one relevant and the other irrelevant. Only the A_1-C and A_2-B_1 pairs showed negative transfer for the relevant pretraining conditions. All three experimental pair types showed negative transfer for the irrelevant pretraining condition.

The generalization interpretation which was presented could not account for the findings. An alternative explanation in terms of extinction or unlearning was suggested.

REFERENCES

BARNES, J. M., & UNDERWOOD, B. J. Fate of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

BESCH, N. F., & REYNOLDS, W. F. Associative interference in verbal paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 554–558.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

MELTON, A. W., & IRWIN, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

PORTER, L. W., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

POSTMAN, L. The present status of interference theory. Paper read at Conference on Verbal Learning, Ardsley-On-Hudson, Nov. 1959.

SPIKER, C. C., & HOLTON, R. B. Associative transfer in motor paired-associate learning as a function of amount of first-task practice. *J. exp. Psychol.*, 1958, **56**, 123–132.

TWEDT, H. M., & UNDERWOOD, B. J. Mixed versus unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58**, 111–116.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

(Received December 27, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 353–360

BRIGHTNESS DISCRIMINATIONS WITH CONSTANT DURATION INTERMITTENT FLASHES[1]

ROBERT L. ERDMANN

IBM Corporation

Under certain conditions intermittent stimulation is more effective than constant stimulation for producing brightness effects. This enhancement is referred to as the Brücke effect and is obtained with low frequency flashing lights. Bartley (1941) has obtained a maximum enhancement with intermittent stimulation rates of about 9 cycles/sec. He has related these brightness enhancement curves to the alpha rhythm of the brain and has postulated that a central neural mechanism is responsible for enhancement.

Using constant duration flashes, Schuil (1940) found a similar brightness effect for frequencies between 5–13 flashes/sec. He found that flashes at these frequencies have an apparent intensity greater than either Blondel-Rey's correction or Talbot's law would indicate. Also using constant duration flashes, Bartlett[2] obtained probability of brightness detection data as a function of flash frequency for rates from 1 to 20 flashes/sec. Although these data are unpublished, the curves showed a maximum detection probability at rates varying from 4 to 12 flashes/sec.

The above studies have all used low frequency intermittent stimulation to obtain a brightness effect. However, they have not been directly concerned with brightness discrimination, or they have confounded flash duration and flash frequency, or there have been special difficulties in specifying the stimulus conditions. At present, systematic data are lacking in which low frequency constant duration flashes have been used as a stimulus for studying brightness discrimination. Therefore, the present study, under conditions which permit reliable specification of the stimulus, was performed to determine the probability of detection of constant duration intermittent flashes as a function of flash frequency with flash and background luminance as parameters.

METHOD

Apparatus.—A diagram of the optical system is shown in Fig. 1. Light for both the adaptation field and the test field was provided by Sylvania R1131C glow tubes (A, A_1). For the adaptation system, light was collimated by an achromatic lens (L_1), passed through a filter holder (F_1), and area limiting field stop (FS_1), a condensing lens (L_2) and formed a Maxwellian view for *S* through a cube beam splitter (BS_2) and 2-mm. artificial pupil. The test field had a similar optical arrangement with the addition of a balanced wedge (BW) to provide a rapid adjustment of luminance.

Light for the fixation pattern was collimated (L_3), transmitted through a red filter (RF), a glass slide (GS), and added to the adaptation system by a beam splitter (BS_1). The fixation pattern consisted of four dots which, edge to edge, subtended a visual angle of 2° in the center of the adaptation field. Fixation luminance was controlled by *S* with a potentiometer and adjusted to a just visible level.

---

[1] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Pure Science, Columbia University. The author wishes to express his appreciation to Conrad G. Mueller for his many helpful suggestions.

[2] Personal communication from Professor N. R. Bartlett on brightness discrimination obtained with a phosphor screen. Unexpected drops in his detection probabilities actually led to the present experiment.

Luminance was calibrated by removing the condensing lens (L_2) from the adaptation field and inserting a diffusing screen and a field stop matched in area (14°30′) and aligned with light from the test field system. Opposite halves of the two fields were blocked and a brightness match made. After turning off the test field and removing the artificial pupil, a Macbeth illuminometer was used to measure luminance. This reading provided a reference luminance in the test field, which, by use of a split field technique, was used to match the luminance of adaptation field when a Maxwellian view was provided.

Once the luminance was calibrated, a constant monitor was available with the aid of a photomultiplier tube (P) built into the apparatus. The photomultiplier circuit, basically that of Hamburger and King (1948), used a 931A photomultiplier tube corrected to approximate the response of the photopic eye by means of a Wratten 15 filter. Thus, any change in the spectral output of the glow tube affected the phototube monitor in the same manner as it would the photic response of the eye. A portion of the light from the cube beam splitter was picked up by the photomultiplier. After amplification, this appeared as a voltage deflection on a Tektronix dual channel scope (Model 545). Matching the voltage deflections with the Macbeth reading provided a monitor of the luminance throughout the experiment.

Figure 2 is a block diagram of the apparatus with associated monitoring equipment. A voltage drop across the glow modulator tube was fed into a Berkeley universal timer and counter (Model 5500) and into one channel of a dual channel Tektronix scope. The Berkeley timer could time the flash duration to the nearest 0.00001 sec. and the interval between flashes with the same accuracy. The timer, along with a calibrated sweep speed on the Tektronix, enabled accurate determination of flash frequency and provided a cross check on the calibration of the two instruments. After the frequency calibrations, the Berkely could be switched over to a counter operation for monitoring the number of flashes in each train of pulses.

To calibrate the wave form of the stimulus, the output of the photomultiplier tube was

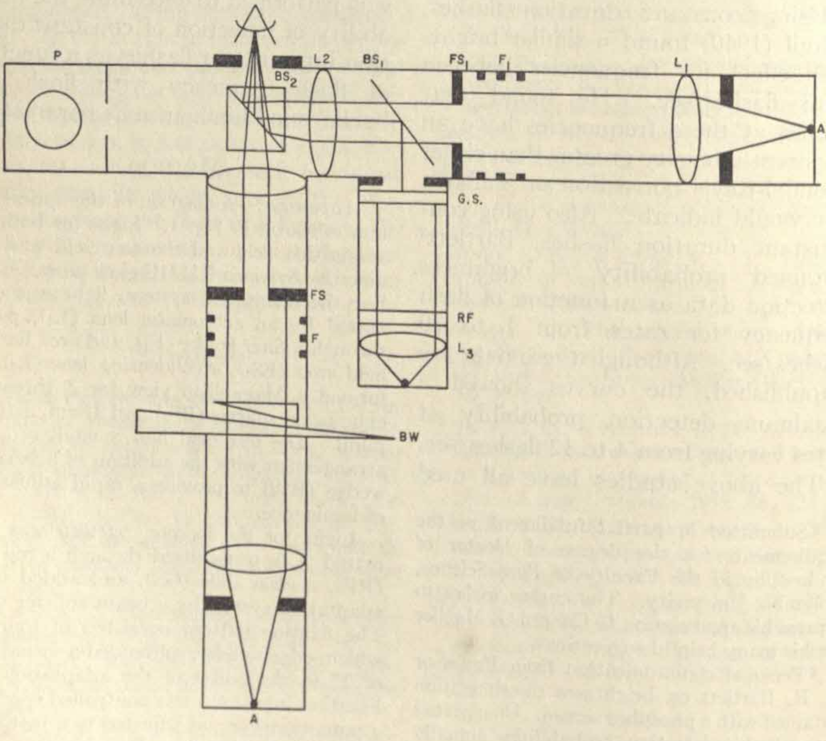


FIG. 1. Schematic of the optical apparatus.

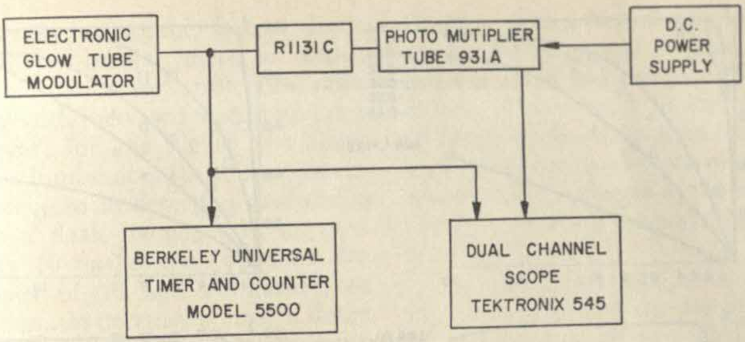

ELECTRONIC GLOW TUBE MODULATOR
R1131C
PHOTO MUTIPLIER TUBE 931A
D.C. POWER SUPPLY
BERKELEY UNIVERSAL TIMER AND COUNTER MODEL 5500
DUAL CHANNEL SCOPE TEKTRONIX 545

FIG. 2. Block diagram of the calibration and monitoring instrumentation.

displayed on the second channel of the Tektronix scope. Since the sweep speed was calibrated, photographing the scope face provided a permanent record of the calibration. In addition, since the voltage drop across the glow modulator tube and the photomultiplier tube output were both displayed on the same scope, their wave forms could be compared. With the exception of a small spike, 6 μsec. in duration, which appeared in both the voltage drop across the glow tube and in the response of photomultiplier, the wave form appeared square.

Prior to each experimental session, all electronic equipment was turned on and allowed to warm up for at least 15 min. After warmup, a calibration check was made on the wave form, duration, and frequency of the test flashes, as well as the light output of the glow modulator tubes. Although the duration of the test flashes was held constant at 1 msec., nine flash frequencies, varying from 1 to 20 flashes/sec, were used in this experiment.

Procedure.—To begin an experimental session *S* would enter a completely darkened cubicle and dark adapted for 15 min. Both *S*s in this experiment were 31-yr-old males having emmetropic vision and experienced as *S*s in visual studies. Dark adaptation was followed by having *S* light adapt to the background luminance for 1 min. The *S*'s position with respect to the eyepiece, which was extended through the wall into the cubicle, was maintained with the aid of a bite board. The background field subtended a visual angle of 14°30′ with the test field subtending 1° of visual angle and appearing in the center of the centrally located fixation field. After the 1 min. of light adaptation, *E* would alert *S* with a buzzer and during the next few seconds present the test stimulus. The *S*'s task was to give one signal with a buzzer when he detected one or more flashes, and two signals with the buzzer when no flashes were detected.

For a given experimental session nine flash frequencies were presented in a randomized order holding flash luminance and background luminance constant. Each frequency was presented 20 times during each session with approximately 20 sec. elapsing between presentations. Three different background luminances and five flash luminances were used, making a total of 15 sessions for each *S*. Five-minute breaks were taken every 15 min., or more frequently at *S*'s request, making an experimental session last about 80 min.

RESULTS

Figure 3 shows the percentage of positive responses as a function of flash frequency for each *S* for three different background luminances, and five flash luminances.[3] Since stimulus presentation was limited to 1 sec., flash frequency also represents the total number of flashes in each test presentation. Each curve represents the data of 1 *S* for one flash luminance at one background luminance. The smooth curves are visual fits with

---

[3] Complete data for Subjects E and J have been deposited with the American Documentation Institute. Order Document No. 7036 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

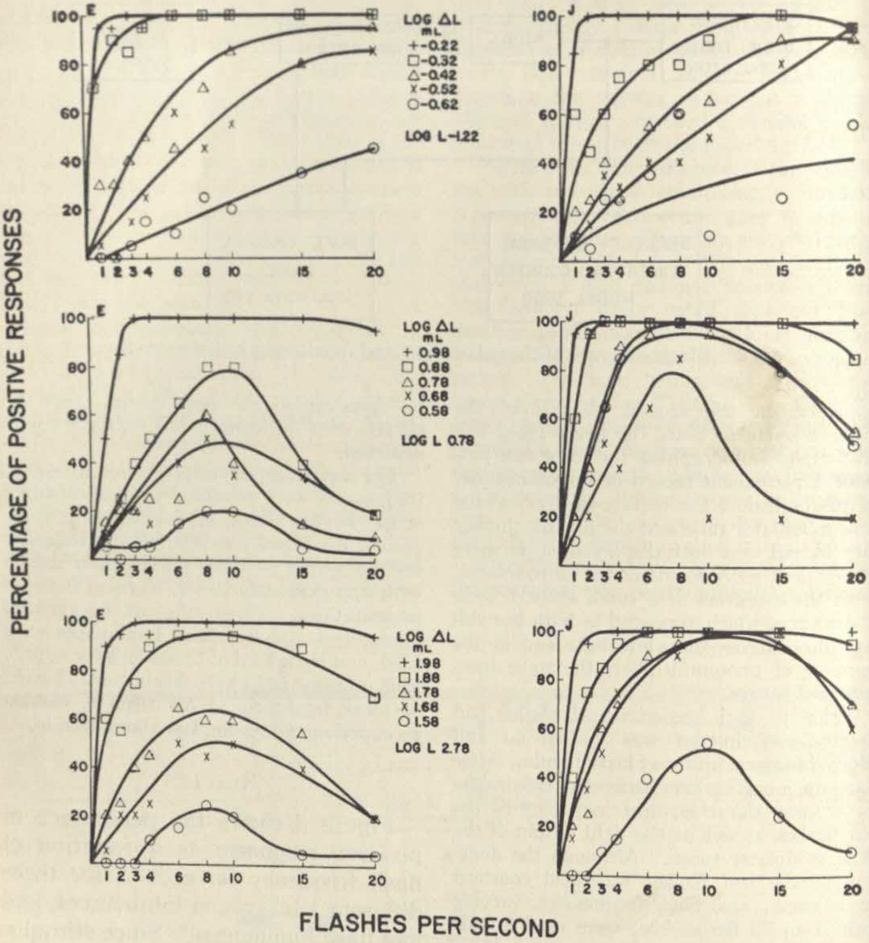


FIG. 3. The percentage of positive responses for 20 presentations as a function of flash frequency for 2 *S*s, five flash luminances, and three background luminances.

the rise part assuming a probability expectation.

Instead of a smooth regular increase in detection probability with increased flash luminance, an irregular increase occurs with an occasional reversal. These irregularities are probably due to the experimental procedure. Since the primary independent variable was flash frequency, the test flash luminances were selected by pilot work before the regular experiment. Once selected, the flash luminance was held constant for each experimental session. Thus, the observed irregularities are within the session to session variability that occurs in gathering brightness discrimination data. In fact, Mueller (1951) has shown that, using the same apparatus and same *S*, a session to session variability of 0.3 log units of test luminance can be expected.

Under every condition, both *S*s showed a systematic increase in detection probability with increasing flash luminance. For the lowest background luminance, there is also a systematic increase in detection probability with increased flash frequency. However, for the higher two background luminances there is a systematic increase in detection probability up to a flash frequency of approximately 10 flashes/sec. At flash frequencies of 15 and 20 flashes/sec, a systematic decrease occurs in detection probability. This decrease occurs even though there is an increase in the number of flashes delivered, which represents an increase in the amount of energy delivered and should therefore represent an increase in the opportunity for detection to occur.

With the intermittent stimulation consisting of a series of constant duration flashes of equal luminance, it is possible to assume that each flash is an independent opportunity for detection and determine the probability of detection on a chance basis. If one assumes that other parameters which might influence the probability are constant, the formulation could take the following form:

$$P = 1 - (1 - p)^n$$

P is the probability of seeing at least one flash.

p is the probability of seeing a single flash in a single flash presentation.

n is the number of flashes presented.

One test of this formulation would be to take the empirical P value for any given n value and compute the other values. However, a limited number of observations were gathered under each condition, and there would be a considerable variation in the expected curves depending upon which specific n was selected.

An alternative test could be obtained by plotting the log of $1 - P$ as a function of n. This should result in a straight line with a slope of $\log 1 - p$. Figure 4 shows the data of *S* E for the lowest background luminance plotted in this manner. Since the log of $1 - P$ becomes indeterminate as a probability of 1.0 is reached, only four curves are shown with a fit obtained by at least squares method.

The data plotted in Fig. 4 do not show any systematic deviation from the linear fit. However, it is evident from Fig. 3 that a systematic deviation would occur with this type probability formulation at flash frequencies of 15 and 20 flashes/sec at the higher background luminances.

To compare more readily the curves shown in Fig. 3 with this probability formulation, Fig. 5 was prepared by taking the values of $1 - p$ obtained in Fig. 4 and computing theoretical curves for P. In this way, all the

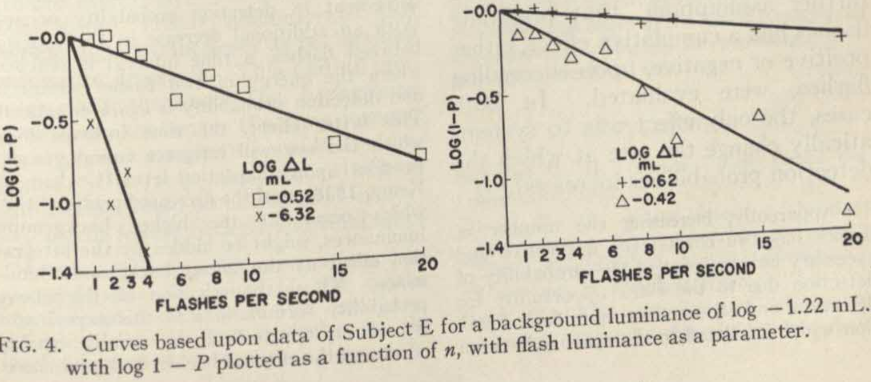


FIG. 4. Curves based upon data of Subject E for a background luminance of log −1.22 mL., with log 1 − P plotted as a function of n, with flash luminance as a parameter.

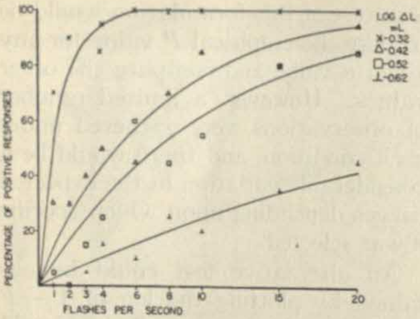


FIG. 5. Curves based upon the data of Subject E for a background luminance of log −1.22 mL., with the percentage of positive responses plotted as a function of flash frequency and flash luminance as the parameter. (The curves show a least squares solution from the data plotted in Fig. 4.)

empirical data contributed to the expected curve. Figure 5 shows the expected and observed values of *P* plotted as a function of *n* with flash luminance the parameter. Both Fig. 4 and 5 tend to support the notion that the increase in detection probability with increasing flash frequency is due primarily to the increased number of chances available to make such a detection.

The data of this experiment do not seem adequate to reject the probability formulation except at the higher luminance levels for frequencies of 15 and 20 flashes/sec. Since these decreased probabilities reject the independence assumption, two further assumptions, that preceding flashes had a cumulative effect, either positive or negative, upon succeeding flashes, were evaluated. In both cases, the only effect was to systematically change the rate at which the detection probabilities increased.[4]

DISCUSSION

Within critical duration, Davy (1952) found that the breaking up of a light stimulus into multiple flashes had no effect upon threshold. Beyond critical duration, threshold energy rises, until, at a separation of 0.5 sec., it is the same for each of two flashes as for one flash and *S*s report two flashes.

If the time interval is great enough, Clark (1958) has shown that a simple probability formulation can account for the increase in detection probability for two flashes over a single flash. The present experiment indicates that the formulation can probably be extended to a somewhat higher number of flashes, particularly if the background luminance is low or the time between flashes is extended.

Perhaps the decreased detection probability is related to the response of a single cell to two light flashes reported by Hartline, Wagner, and MacNichol (1952). This work demonstrated that a flash of light too weak to elicit a response had some excitatory effect that persisted for more than 2 sec. The excitatory effect was demonstrated by using a second flash to test the sensitivity of the receptor at various times after exposure to a subliminal flash. Following a period of enhanced sensitivity to the second flash there was a period of diminished sensitivity followed by a return to a normal excitatory state.

---

[4] Apparently, increasing the number of flashes has two effects: (*a*) a positive and necessary enhancement of the probability of detection due to the extra opportunity for detection, and (*b*) an effect upon flash detection which is related to the spacing between flashes. If the spacing between flashes is great enough, the flashes represent independent opportunities for detection. As the spacing between flashes decreases, a time between flashes interval is reached where a decrement in detection probability occurs. With an additional decrease in the spacing between flashes, a time interval is reached where the energy of two flashes summate and detection probability is again enhanced. This latter effect, the time interval over which the eye will integrate energy, is dependent upon adaptation level (Graham & Kemp, 1938), and the decreased probabilities, which occur at the higher background luminances, might be hidden by the integration effect at the lowest background luminance. Thus, the rejection of the above probability formulations in this experiment is probably due to both the variable spacing effect and the effect of background luminance.

Similarly, in a train of flashes, the same relative time sensitivities might exist. When flash frequency is low, the receptor unit has time to return to its normal excitatory state and each flash represents an independent opportunity for detection. As flash frequency is increased, a time interval between flashes is reached corresponding to the period of diminished sensitivity. Thus, if the flashes in a train are timed so that each flash after the first arrives during this period of diminished sensitivity, then one might expect a decrease in the probability of detection. With a further increase in flash frequency, a time period is reached corresponding to enhanced sensitivity and one flash contributes to the detection of a following flash.

Within critical duration, Long (1951) showed the wave form of energy distribution to be independent of its effect, and Davy (1952) indicated that the number of flashes times luminance equals a constant. These experiments demonstrate that a frequency or time interval is reached where integration of energy must occur. Further increases in frequency of constant duration pulses simply present more energy to the eye and lead to an apparent brightening of the test field.[5]

If a period of diminished sensitivity accounts for the observed decreased detection, what accounts for the failure to observe this effect at the lowest background brightness? The data of Graham and Kemp (1938) demonstrate that critical duration increases as background luminance decreases. This finding indicates that the lower the background luminance, the longer the time interval during which two flashes would be integrated. In fact, their data indicate summation would occur above 14 cycles/sec for the lowest background luminance used in this experiment and above 20 and 30 cycles/sec for the two higher luminance levels. Thus, one might expect the diminished sensitivity, which exists for flash frequencies of 15 and 20 flashes/sec at the higher background luminances, to be hidden, replaced, or counteracted by the temporal summation effect at the lowest background luminance.

SUMMARY

The probability of detection of constant duration intermittent flashes was studied as a function of flash frequency with flash luminance and background luminance as parameters. The *S*s were adapted to a 14°30′ field of constant luminance and presented with a 1° test flash train in the center of the field. The duration of the stimulus train of flashes was 1 sec. Twenty observations were gathered on 2 practiced *S*s for each of nine frequencies ranging from 1 to 20 flashes/sec.

The results indicated: (*a*) increasing flash luminance increases the probability of detection for every condition; (*b*) for the condition of low background luminance, probability of detection increases as flash frequency increases; and (*c*) for the higher background luminances, there is a systematic increase in detection up to flash frequencies of approximately 10 flashes/sec, followed by a systematic decrease in probability of detection for frequencies of 15 and 20 flashes/sec.

The increase in the number of chances for flash detection with increasing flash frequency may account for the initial increase in probability of detection that occurs for all luminance values. The decreases in detection probability for the higher flash frequencies at the higher background luminance levels can possibly be attributed to a period of diminished sensitivity that is dependent upon the time interval between flashes. The failure to observe this effect at the lowest background luminance can possibly be attributed to a difference in critical duration.

REFERENCES

BARTLEY, S. H. *Vision.* New York: Van Nostrand, 1941.

CLARK, W. C. Visual detection threshold as a function of pulse duration and separation. Unpublished doctoral dissertation, University of Michigan, 1958.

DAVY, E. The intensity-time relation for multiple flashes of light in the peripheral

---

[5] Although only a few observations were gathered after conclusion of the regular experimental session, *S*s of this experiment indicate that this predicted effect is the observed effect.

retinal. *J. Opt. Soc. Amer.*, 1952, **42,** 937–941.

GRAHAM, C. H., & KEMP, E. H. Brightness discrimination as a function of the duration of the increment in intensity. *J. gen. Physiol.*, 1938, **21,** 635–650.

HAMBURGER, F., JR., & KING, E. J. Recording photometer and its use in studies of cathode-ray screen displays. *J. Opt. Soc. Amer.*, 1948, **38,** 875–879.

HARTLINE, H. K., WAGNER, H. G., & MACNICHOL, E. F., JR. The peripheral origin of nervous activity in the visual system. *Symp. Quant. Biol.*, 1952, **17,** 125–141.

LONG, G. E. The effect of duration of onset and cessation of light flash on the intensity-time relation in the peripheral retina. *J. Opt. Soc. Amer.*, 1951, **41,** 743–747.

MUELLER, C. G. Frequency of seeing functions for intensity discriminations at various levels of adapting intensity. *J. gen. Physiol.*, 1951, **34,** 463–474.

SCHUIL, A. E. The fixed light equivalent of flashing lights having constant duration. *Trans. Illum. Engng. Soc., Lond.*, 1940, **5,** 117.

(Received January 3, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 361–369

THE TEMPORAL COURSE OF PROACTIVE INHIBITION FOR SERIAL LISTS[1]

LEO POSTMAN

University of California

Application of interference theory to the analysis of long-term retention has brought into focus the importance of proactive inhibition (PI) as a mechanism of forgetting. As Underwood (1957) has shown, classical studies of retention using practiced *S*s in counterbalanced designs greatly overestimated the amount of forgetting because of substantial amounts of PI accumulated during the learning of successive lists. Retention of single lists by naive *S*s is considerably higher than had been found with practiced *S*s in earlier investigations (Postman, 1961; Postman & Rau, 1957; Underwood & Postman, 1960; Underwood & Richardson, 1956). The retention losses observed with naive *S*s may reasonably be attributed to retroactive inhibition (RI) and PI from verbal habits practiced outside the laboratory. More of these habits are likely to have been acquired prior to the experiment than during the retention interval. Hence extraexperimental interferences should be largely proactive rather than retroactive.

The conditions and characteristics of extraexperimental interference are assumed to be the same as those observed in formal studies of RI and PI. Specifically, the acquisition of the experimental list is believed to involve the unlearning or extinction of pre-experimental associations which *S* brings to the laboratory. The extinguished associations recover during the retention interval and compete with the prescribed responses at the time of recall. The amounts of extinction and recovery depend, of course, on the number and strength of the pre-experimental associations in competition with the prescribed responses. Thus, there should be more competition from extraexperimental sources for words of high than low meaningfulness. The patterns of errors in learning and recall obtained with materials of high and low meaningfulness are consistent with this expectation (Postman, 1961; Underwood & Postman, 1960).

A central assumption of the present analysis of forgetting is that PI increases as a function of time. This assumption may be considered correct by definition. In experiments on interserial interference, PI is obtained even if the learning of the second list is carried to a point at which associative inhibition is no longer present or is superseded by associative facilitation. The necessary condition for obtaining PI is that the retention interval be longer than the intertrial interval in the acquisition of the second list. If the process responsible for the interference is continuous, PI must increase as a function of time (Underwood, 1945, 1948a). It is important to ask, however, whether the rate of increase in PI makes it reasonable to attribute long-term forgetting largely to proactive interference from extraexperimental sources.

Experimental evidence on the temporal course of PI for verbal habits is not extensive. One study explicitly concerned with this problem is that of Underwood (1949) which used paired adjectives, with practiced *S*s serving in a counterbalanced design. No increase in PI was found between 20 min. and 75 min. after the end of learning. There was, however, a substantial increase in interlist intrusions, indicating a loss of differentiation as a function of time. In

---

[1] This research was supported by a grant from the National Science Foundation.

another experiment using paired adjectives, Underwood (1948a) compared RI and PI after 5 and 48 hr. After 5 hr. RI was significantly greater than PI. This difference was no longer present after 48 hr.; retention of the second list had declined whereas there was a slight increase in the recall of the first list. Since PI is a function of the degree of prior learning, the difference in the rate of forgetting for the two lists suggests that there should be long-term increases in PI. The results of experiments using the method of modified free recall (MFR) lead to the same prediction since they show long-term increases in the frequency of responses from the first list along with a steady decline in responses from the second list (Briggs, 1954; Briggs, Thompson, & Brogden, 1954; Underwood, 1948b).

This experiment investigates the temporal course of PI for serial lists with naive *Ss*. As in some of the recent studies of extraexperimental interference (Postman, 1961; Underwood & Postman, 1960), words of high (HF) and low frequency (LF) of usage were used. It was expected that (*a*) PI at recall would increase significantly during a 1-week retention interval, and (*b*) the loss of differentiation, as measured by interlist intrusions, would be greater with materials of high than low frequency. The latter expectation was based on the assumption that there would be greater overlap in the associative context of successive lists with materials of high frequency.

METHOD

Design.—The three independent variables were Condition of Practice (Experimental vs. Control), Word Frequency (High vs. Low), and Retention Interval (30 sec., 30 min., and 7 days). The Experimental groups learned three successive lists and were tested for recall of the third. The Control groups learned and recalled a single list. The results of the 30-sec. test were used to assess associative strength at the end of learning, in view of expected differences in rate of acquisition as a function of word frequency and interlist transfer. With two conditions of practice, two levels of word frequency, and three retention intervals, the total design comprised 12 independent groups.

Lists.—The learning materials were lists of 12 one-syllable nouns. There were three HF lists and three LF lists. For the HF words the number of occurrences in the "L" count of Thorndike and Lorge (1944) was between 1000 and 3300 in 4.5 million; for the LF words the number of occurrences ranged between 1 and 3. There was no repetition of first letters in any of the lists. Six different serial orders were used equally often with each of the lists. The orders were so constructed that no word appeared more than once in any serial position and that no forward sequence of two words was used more than once.

There were no synonyms among the words of a given frequency. At each frequency level, the three lists had 10 first letters in common; 2 of the first letters in each list were not repeated in that position in the other lists. For the Experimental groups, the six possible orders of the three lists were used equally often. For the Control groups each of the single lists was used equally often. The lists are presented in Table 1.

Procedure.—The lists were presented on a Hull-type memory drum at a 2-sec. rate, with an intertrial interval of 6 sec., and learned by the anticipation method to a criterion of one perfect trial. In the experimental condition, successive lists were separated by an interval of 1 min. Retention was tested by relearning for five trials or to a criterion of one perfect trial, whichever took the longer. For the groups tested after 30 min., the retention

TABLE 1

LISTS USED IN THE EXPERIMENT

High Frequency			Low Frequency		
List 1	List 2	List 3	List 1	List 2	List 3
BED	BOY	ARM	BOLE	BARD	BRIG
CHILD	CHAIR	BILL	CARP	CRYPT	CAW
DRESS	DROP	CAR	FINCH	FLUME	DRAM
FLOOR	EYE	DOOR	GIG	GUILE	FOUNT
JOB	FRONT	FIGHT	JAG	LOUT	GROAT
KISS	LINE	MIND	LEEK	MEIN	IRE
MILE	MONTH	NEED	MIDGE	PUG	LODE
NAME	NIGHT	PRICE	PRAWN	QUOIT	MALL
PART	POST	SIDE	SKEIN	SKIFF	PRISM
SENSE	STATE	TRAIN	TUSK	TALC	SLAG
TOUCH	TOWN	VOICE	VANE	WART	TITHE
WALL	WIFE	WORLD	WOMB	YAP	WENCH

Note.—Six different random orders were used with each list.

interval was filled with two tasks—mathematical problems (20 min.) and rating of pictures (10 min.).

Subjects.—There were 12 groups of 12 *Ss* each. The *Ss* were assigned to conditions in blocks of 12, with 1 *S* per block from each of the groups of the experimental design. The running order within each block was determined by a table of random numbers, as was the assignment of lists and serial orders within each condition. All *Ss* were undergraduate students who had English as their native language and were naive to rote-learning experiments. No *Ss* had to be rejected because of failure to learn.

RESULTS

First-list learning.—For List 1—the first of the two prior lists of the Experimental groups and the test list of the Control groups—there were no significant differences among the groups learning material of a given frequency. The results of the 72 *Ss* learning each type of list have, therefore, been combined. The mean number of trials to criterion for the HF lists was 16.83 (SD = 6.07), and 21.14 (SD = 8.02) for the LF lists. The difference in speed of learning is significant beyond the .001 level ($t = 3.59$, $df = 142$).

The mean number of misplaced responses per trial was 1.46 (SD=.63) for the HF lists and 1.22 (SD = .51) for the LF lists. This difference is highly significant ($t = 8.00$, $df = 142$, $P < .001$). Misplaced responses account for all but a few of the overt errors during acquisition. The mean percentages that misplaced responses were of all overt errors are 98.4 and 95.1 for the HF and LF lists, respectively. The difference in favor of the HF groups is maintained when the misplaced responses are expressed as percentages of opportunities (total occurrences minus correct responses). The mean percentage is 28.4 (SD = 13.9) for the HF lists, and 22.4 (SD = 9.7) for the LF lists. The difference in percentages (following arc-sine transformation) is also significant ($t = 2.65$, $P < .01$). As in an earlier study using two-syllable nouns from the same frequency ranges, speed of learning and error rate are directly related (Postman, 1961).

Transfer effects in original learning.—The mean numbers of trials to criterion for the three successive lists learned by the Experimental groups are shown in Table 2. For both kinds of materials, there is a substantial increase in speed of learning on List 2 and little further change on List 3. The relative amounts of improvement are comparable for the HF and LF lists. The effects of order of learning are significant (F=112.62, df=2/140,

TABLE 2

TRIALS TO CRITERION AND RESULTS OF STAGE ANALYSES FOR SUCCESSIVE LISTS LEARNED BY THE EXPERIMENTAL GROUPS

List	Word Frequency			
	High		Low	
	Mean	SD	Mean	SD
	Trials to Criterion			
1	17.25	6.08	21.31	6.95
2	10.06	3.73	11.67	4.70
3	8.69	3.65	11.78	5.48
	Trials for Response Learning[a]			
1	4.24	1.72	5.24	1.36
2	2.97	.82	3.99	1.14
3	2.89	.96	3.89	.85
	Trials for Associative Learning[b]			
1	1.39	1.01	1.69	1.62
2	.40	.47	.34	.38
3	.29	.34	.25	.35

[a] Mean trial on which response was given anywhere in the list.

[b] Mean number of trials between first occurrence and first correct placement of response.

$P < .001$) but do not interact with frequency ($F = 1.67$).

Since there was substantial "learning to learn," the test was mastered considerably faster by the Experimental groups than the Control groups. Differences in learning are likely to influence retention, and it becomes important to inquire what characteristics of performance were responsible for the positive transfer effects. We considered first the effects of practice on the two stages into which the period of acquisition can be divided analytically—the response-learning stage and the associative stage (Underwood & Schulz, 1960, pp. 92–94). The duration of the response-learning stage is estimated by the mean trial on which the correct responses are given anywhere in the list. The mean number of trials between the first occurrence and the first correct placement of the responses measures the length of the associative stage. Table 2 shows the results of stage analyses for the three successive lists learned by the Experimental groups. The duration of the response-learning stage declines sharply for both kinds of materials between List 1 and List 2 but does not drop further on List 3. The variation among lists is significant ($F = 44.44$, $df = 2/140$, $P < .001$) but order of learning does not interact with frequency ($F < 1$). Throughout practice, the response-learning stage remains shorter for the HF than the LF lists. The over-all effect of frequency is significant beyond the .001 level ($F = 21.90$, $df = 1/70$).

The changes in the duration of the associative stage parallel those for the response-learning stage. Again there is a clear increase in speed between Lists 1 and 2 and no further change on List 3. Order of learning is significant ($F = 17.84$, $df = 2/140$, $P < .001$) whereas the interaction of order with frequency is not ($F < 1$). In contrast to the results obtained for the response-learning stage, the differences between the two kinds of materials are small and inconsistent. The over-all effect of frequency is not significant ($F < 1$). It may be noted that the same conclusions concerning the effects of frequency on the two stages of learning are reached when the performance of all *S*s on List 1 is considered. The difference between the HF and LF conditions is significant for the response-learning stage ($t = 4.52$, $df = 142$, $P < .001$), but not for the associative stage ($t = 1.55$).

A second systematic effect of practice is the reduction in the relative frequency of misplaced responses. Table 3 presents for Lists 1–3 the mean numbers of misplaced responses per trial as well as the mean percentages of such errors based on the number of opportunities. The mean number of errors per trial declines from list to list whereas there is a clear drop in the percentages only on List 3. A decline in the number of errors per trial may merely be complementary to an increase in correct responses, and the percentages probably provide the clearest index of *S*s' disposition to give misplaced responses. In any event the rate of misplaced responses is clearly lower for List 3 than for List 1. The change in both indices is significant beyond the .01 level ($F = 5.36$ for responses per trial and 7.72 for percentages, $df = 2/70$ in both cases). In neither case is there a significant interaction between order and frequency ($F = .06$ and 1.78, respectively).

Interlist intrusions during learning.—The number of interlist intrusions during the acquisition of Lists

TABLE 3

MEAN NUMBERS PER TRIAL AND PERCENTAGES OF MISPLACED RESPONSES IN ACQUISITION OF SUCCESSIVE LISTS BY EXPERIMENTAL GROUPS

List	Word Frequency					
	High			Low		
	Mean	*SD*	%	Mean	*SD*	%
1	1.40	.57	26.8	1.20	.52	22.0
2	1.00	.45	26.2	.91	.62	19.7
3	.84	.46	23.0	.68	.47	15.3

2 and 3 was small. Under the HF condition, a total of 5 intrusions was contributed by 3 *S*s during the learning of List 2, and a total of 10 intrusions by 7 *S*s during the learning of List 3. All of the latter came from List 2. Under the LF condition, there were 4 intrusions obtained from 3 *S*s during the learning of List 2, and 9 intrusions from 7 *S*s during the learning of List 3. Of these 9 intrusions, 6 came from List 1 and 3 from List 2. There is a suggestion here that the extinction of List 1 responses during acquisition of List 2 was less effective with materials of low than high frequency.

Recall.—The mean numbers of items recalled by the various groups are shown in Fig. 1. In view of the differences in rate of acquisition of the test list as a function of both frequency and interlist transfer, the results of the 30-sec. tests are of interest. Under the control condition, the postcriterial drop is, as expected, greater for the LF list than the HF list. Under the experimental condition, however, the reverse is true in spite of the fact that the HF list was learned faster than the LF list. This reversal reflects the fact that with HF materials the drop is greater for the Experimental than the Control group whereas with LF materials the difference is in the opposite direction. For the results of the 30-sec. tests, the interaction of condition with frequency approaches, but does not reach, significance ($F = 3.84$, whereas 4.08 is required at the .05 level with 1/44 *df*). The pattern of differences suggests that proactive effects may develop extremely rapidly with HF materials and may influence even a retention test immediately after the end of learning. The fact that there were no interlist intrusions on the 30-sec. tests does not contradict this interpretation since differentiation of the lists should be high at this point. In any event, it is important to note that the usual inverse relationship between speed of learning and postcriterial drop does not necessarily hold true under conditions of transfer.

We turn now to the results of the delayed tests. Under the control condition the rate of forgetting for the two kinds of materials is comparable. During the 1-week interval (30 sec.–7 days) the mean retention

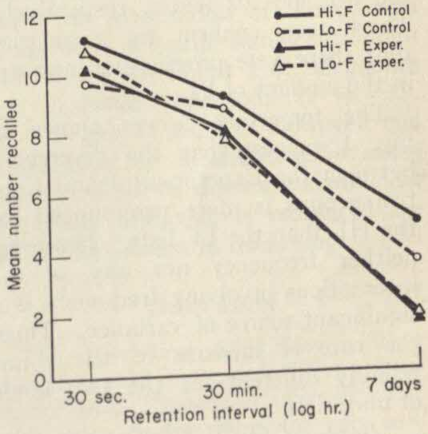


FIG. 1. Mean numbers of items correctly recalled by the Experimental and Control groups as a function of retention interval.

loss is 6.25 for the HF lists, and 6.17 for the LF lists. The Experimental groups recall less than the Control groups, and this difference increases for both types of lists between 30 min. and 7 days. Following a Freeman-Tukey square-root transformation, the recall scores were subjected to an analysis of variance. The differences among retention intervals are significant beyond the .001 level ($F = 137.21$, $df = 1/132$) as is the difference between Conditions ($F = 13.55$, $df = 1/132$). The interaction, Time × Conditions, is also significant ($F = 7.25$, $df = 2/132$, $P < .01$) and reflects the increasing difference between Experimental and Control groups as a function of time. In order to determine whether there was a reliable increase in PI between 30 min. and 7 days, the interaction variance was partitioned into two components with 1 *df* each—that associated with the differences between the 30-sec. test and the combined delayed tests on the one hand, and that associated with the differences between the two delayed tests on the other. The two *F* ratios of 7.73 and 6.78 are significant beyond the .01 and .02 levels, respectively. These tests confirm the conclusion that there were progressive increases in the amount of PI.

The forgetting curves shown in Fig. 1 suggest that the divergence between the Experimental and Control groups is more pronounced for the HF than the LF lists. However, neither frequency nor any of the interactions involving frequency is a significant source of variance. Thus, the rate of increase in PI is not reliably different for the two kinds of materials.

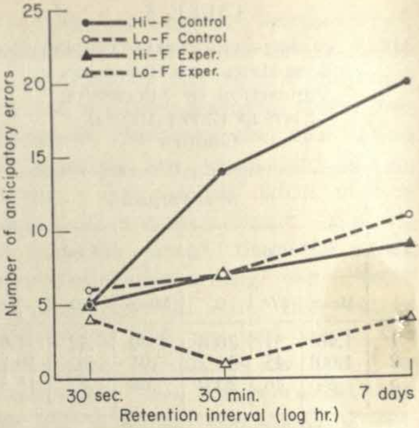


FIG. 2. Numbers of different anticipatory errors at recall.

Misplaced responses at recall.—The number of different anticipatory errors was used to measure the tendency to misplace correct responses at recall. Each such error represents a response which is available to *S* before re-exposure to the item but whose serial position has been forgotten. As Fig. 2 shows, the degree of recovery of these errors corresponds to the rate of misplaced responses during acquisition of the test list. Thus, the errors recover more rapidly for the Control groups than the Experimental groups, and for the HF lists than the LF lists. In view of the large number of zero scores, the following procedure was used to evaluate the significance of the temporal trends. Each *S* in the 30-min. and 7-day groups was matched with the *S* in the appropriate 30-sec. group in the same block. The difference in the number of errors was determined for each pair, and the distributions of difference scores were subjected to an analysis of variance. The scores of the 2 *S*s matched to the same 30-sec. *S* were treated as repeated measurements. The variations as a function of frequency and condition are significant beyond the .05 level ($F = 4.25$, $df = 1/44$ in both cases since the mean differences were the same for the two effects). The difference between 30 min. and 7 days is not significant ($F = 2.55$, $df = 1/44$), and there are no significant interactions. The relatively small increases in anticipatory errors between 30 min. and 7 days probably reflect reduced availability of responses from within the list. The increases are larger when total numbers of errors rather than numbers of different errors are considered since *S*s with low recall scores tend to repeat the same errors several times in succession.

Interlist intrusions at recall.—Table 4 shows the frequencies of interlist intrusions at recall and the numbers of *S*s contributing them. There were no intrusions on the 30-sec. tests. The total numbers of intrusions on the delayed tests are small but there are two consistent trends. First, there is an increase in intrusions as a function of time which may be attributed to a loss of list differentiation. Second, more intrusions are given during the recall of the HF lists than the LF lists, indicating that the loss of differentiation is greater for the former material. Finally, most of the intrusions are from the second of the two prior lists. This finding supports Underwood's (1945) conclusion that contiguity during learning is conducive to interlist confusion.

TABLE 4

NUMBERS OF INTERLIST INTRUSIONS AT RECALL

Word Frequency	Retention Interval			
	30 Min.		7 Days	
	Total	N^a	Total	N^a
High	4(4)	2	8(6)	5
Low	—	—	3(3)	2

Note.—Numbers from second prior list are in parentheses.
[a] Number of *S*s giving intrusions.

TABLE 5

MEAN NUMBERS OF CORRECT ANTICIPATIONS IN FIVE TRIALS OF RELEARNING

Retention Interval	Cond.	Word Frequency			
		High		Low	
		Mean	*SD*	Mean	*SD*
30 sec.	Exp.	53.42	4.25	53.83	6.03
	Cont.	55.25	2.98	53.33	5.87
30 min.	Exp.	51.75	6.18	47.50	9.84
	Cont.	51.83	5.06	50.42	7.11
7 days	Exp.	33.25	4.99	33.17	10.08
	Cont.	41.08	8.69	39.92	7.49

Relearning.—The mean numbers of correct responses in five trials of relearning are shown in Table 5. The differences between the Experimental and Control groups are quite small on the 30-min. tests but remain substantial on the 7-day tests. The interaction of Time with Conditions is significant ($F = 3.80$, $df = 2/132$, $.02 < P < .05$, after square-root transformation). Thus, the differences in PI as a function of time are not limited to the recall trial. This conclusion is confirmed by the fact that the interaction of Time with Conditions remains significant when the scores on Trial 2 of relearning are analyzed separately ($F = 6.35$, $df = 2/132$, $P < .01$). However, the dissipation of interference is sufficiently rapid to eliminate significant PI when retention is measured by trials to criterion in relearning.

DISCUSSION

The experiment yields clear evidence for substantial increases in PI as a function of time. These results were obtained with naive *S*s for whom the only controlled source of interference consisted of two lists learned immediately prior to the test list. In the light of these findings the assumption that PI from extraexperimental sources increases progressively in time remains reasonable. According to an interference interpretation, both the Experimental and Control groups in the present study were subject to PI, and possibly RI, from extraexperimental sources. For the Experimental groups, additional interference came from the lists learned prior to the test list. The prior lists added substantially to the total effective interference at recall since they were associated with the same experimental situation as the test list. Thus, retention decreased progressively for both the Experimental and Control groups but significantly more so for the former.

In assessing the amounts of PI observed in the present study, it is important to consider the differences between the Experimental and Control groups in the acquisition of the test list. As a result of practice on the prior lists, both the response-learning and associative stages of acquisition were shortened for the Experimental groups, the probability of misplaced responses was reduced, and the speed of mastery to criterion was increased. These effects of "learning-to-learn" should produce greater associative strength at criterion under the experimental than the control condition. The results of the 30-sec. tests cannot take full account of all the differences in learning. For example, since prior learning reduced the rate of misplaced responses, there were fewer such errors which could recover for the Experimental than the Control groups. Such differences in error tendencies first become apparent on the delayed tests of retention (Fig. 2). Thus, large amounts of PI developed during the retention interval in spite of the fact that the conditions of acquisition probably were more favorable to retention under the experimental condition.

There were no significant differences between the amounts of PI obtained with the HF and LF lists. This finding parallels the results obtained with single lists: there was no difference in the rate of forgetting for the two kinds of materials either under the control condition of the present experiment or in an earlier study investigating the retention of serial lists (Postman, 1961). There is a strong indication that under the experimental condition the loss of list differentiation was, as expected, greater with the HF than the LF materials. Although interlist intrusions were given by only part of the *S*s, they were consistently more numerous in the recall of the HF lists.

As in other studies of interserial interference, the number of interlist intrusions at recall was small and accounts for only a fraction of the total retention decrements under the experimental condition. The number of intrusions increases as a function of time but at a much slower rate than total PI. On the assumption that the differentiation of lists learned to criterion is lost slowly and is reinstated readily during the recall trial, at least some of the retention loss can be attributed to implicit intrusions which were rejected by *S*. The assumption that differentiation remained high is indirectly supported by the fact that importations from outside the list were extremely rare under any of the conditions of the experiment. Thus, there was highly effective discrimination between the experimental lists and extra-experimental responses as well as among the successive lists.

SUMMARY

This experiment investigates the temporal course of PI for serial lists. Lists composed of either high-frequency words or low-frequency words were used. The Experimental groups learned three successive lists and were tested for recall of the third. The Control groups learned and recalled a single list. Retention was tested by relearning after intervals of 30 sec., 30 min., and 7 days. There were significant increases in the amount of PI as a function of time for both kinds of materials. There were no significant differences as a function of word frequency, either in the retention of single lists or in the amount and rate of increase in PI. Interlist intrusions increased as a function of time and were more numerous in the recall of the high-frequency than the low-frequency lists. The fact that

PI increases substantially over time is consistent with the hypothesis that the growth of proactive interferences from extraexperimental sources is responsible for much of long-term forgetting.

REFERENCES

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

BRIGGS, G. E., THOMPSON, R. F., & BROGDEN, W. J. Retention functions in reproductive inhibition. *J. exp. Psychol.*, 1954, **48**, 419–428.

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, **61**, 97–110.

POSTMAN, L., & RAU, L. Retention as a function of the method of measurement. *U. Calif. Publ. Psychol.*, 1957, **8**, 271–396.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 273).

UNDERWOOD, B. J. Proactive and retroactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38**, 29–38. (a)

UNDERWOOD, B. J. "Spontaneous" recovery of verbal associations. *J. exp. Psychol.*, 1948, **38**, 429–439. (b)

UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

UNDERWOOD, B. J., & RICHARDSON, J. The influence of meaningfulness, intralist similarity, and serial position on retention. *J. exp. Psychol.*, 1956, **52**, 119–126.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received January 10, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 370–375

TRANSFER OF TRAINING AS A FUNCTION OF TASK DIFFICULTY IN A COMPLEX CONTROL SITUATION[1]

DONALD A. GOLDSTEIN AND JOHN M. NEWTON[2]

Electric Boat Division, General Dynamics Corporation

An important problem in training is that of assessing the effects of initial task difficulty on transfer, i.e., whether training on a more difficult task results in better transfer to a less difficult task or vice versa. Studies attempting to answer this question have produced equivocal results. For example, Ritchie and Michael (1955) have found better transfer from more to less difficult tasks while Lordahl and Archer (1958) have found the reverse. Day (1956), in his review of the area, has called attention to a major problem in generalizing about results: the lack of specificity as to how difficulty was varied.

One area in which it appears possible to specify task difficulty is that of complex control systems, i.e., those systems in which one or more integrations, with corresponding time lags for these integrations, intervene between control action and system reaction. To a large extent, differences in ease of control (task difficulty) reflect both the complexity of the control dynamics involved and the length of the time phase relationship between control and display movements. For longer control lags, *S* must attend more closely to the subtle information (rate and acceleration) than he would have to for shorter control lags, given the same number of integrations. In extreme cases, because of the long delay, *S* may be forced to make compensatory control movements before the effects of his previous movement(s) become apparent on the display. Thus, differences in control lag may reflect differences in ease of control (task difficulty). Differences in control lag, however, are themselves a function of certain physical restraints inherent in the system, such as moment of inertia, system gain, etc., and some of these, in turn, can be identified with certain coefficients in the dynamic equations of motion representing the system. By specifying task difficulty in terms of changes in the values of certain coefficients in the dynamic equations of a complex control system, it should be possible to determine experimentally: (*a*) which resulting system, of those chosen to study, proves to be most difficult to control, and (*b*) which system produces, on the average, the greater positive transfer to all other systems studied.

To investigate these effects, coefficients affecting moment of inertia and speed of system reaction in the dynamic equations of motion of a submarine (a third-order integration system) were varied, each variable taking two separate sets of values. In combination, this procedure defined four separate training conditions. Through the use of a factorial design all possible transfer combinations among the four conditions were observed.

---

[1] This research was supported by the United States Navy under Contract NOnr 3075(00), monitored by the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Washington 25, D. C. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government. The authors wish to express their appreciation to Lawrence Hill and George Simoneau of the Human Factors Staff for their aid in the design and construction of the apparatus used in this study.

[2] Now at the University of Omaha, Omaha, Nebraska.

METHOD

Apparatus.—The control-display system used in this study was the same for all *S*s. The display consisted of a straight scale indicator, 18 in. long, calibrated from ±150 ft. in 10-ft. increments, with ½ in. equal to 10 ft. The *S*'s task was to keep a moving pointer, ½ in. long, at a zero reading on the scale. The electrical limits of the system were ±100 ft., with 10 v. equal to 100 ft. The error generator employed a sine wave forcing function to move the pointer within a range of ±50 ft. at the rate of .006 cps with an amplitude of ±2.5 v. The position on the sine wave at the start of the trial was randomized across trials.

The control, a modified joystick, 6¾ in. long and 1 in. diameter, had a maximum deflection angle of ±45° with the stick voltage output at full travel equal to ±75 v. The equivalent plane angle at full travel was ±25°, making the angular deflection of the plane roughly equal to one-half the angular deflection of the control stick.

A Heathkit 15 amplifier computer was utilized with 10 analog amplifiers used for the submarine programs and 4 amplifiers used for the scoring circuits. A Standard Electric timer was used to record time on a target zone of ±10 ft. for the last 100 sec. of each 120-sec. trial. The programs employed were simplified equations of submarine motion in the *Z* axis, with differing constants depending upon the moment of inertia and speed of reaction of the submarine being simulated.

Table 1 shows the simplified equations of motion used in the computer program. These equations are a result of maintaning a constant speed within any trial. Values of the constants are not presented since they are completely artificial.

Subjects.—The *S*s were 80 undergraduate students, 58 men and 22 women, enrolled at Mitchell College in New London, Connecticut. An additional 5 *S*s had to be eliminated from the study for failure to follow directions or as a result of equipment breakdown. All *S*s were volunteers taken from undergraduate courses and they were paid $1.00 per hour for their participation.

TABLE 1

SIMPLIFIED EQUATIONS OF MOTION USED IN THE COMPUTER PROGRAM

Type of Motion	Equation
For pitch:	$\ddot{\theta} = -k_1\dot{\theta} - k_2(k_{11}\theta - k_3W) \pm k_4\delta_s$
For normal force:	$\dot{W} = k_5\delta_s + k_6\dot{\theta} - k_7W$
For depth rate:	$\dot{Z} = k_8\theta - k_9W$
For depth:	$Z = k_{10}\dot{Z}$

where: θ = displacement in the horizontal attitude
Z = displacement in the vertical attitude
U = axial velocity and is contained within the k values
W = virtual mass
δ_s = displacement of the control stick
.. = acceleration terms
. = rate terms

Design.—The factorial design for this study employed four dimensions: moment of inertia trained on, moment of inertia tested on, speed of system reaction trained on, and speed of system reaction tested on. As each of these dimensions had two levels, large and small moment of inertia, fast and slow speed of system reaction, the final analysis resulted in 16 cells (2 × 2 × 2 × 2), with five replications within each cell.

Procedure.—The 80 *S*s were randomly assigned to 1 of 16 groups, with 4 groups training under each of the original conditions. These conditions represented the effects of a small moment of inertia with a fast speed of system reaction (SF), a small moment of inertia with a slow speed of system reaction (SS), a large moment of inertia with a fast speed of system reaction (LF), and a large moment of inertia with a slow speed of system reaction (LS). After training, each of the 4 groups for each original condition was transferred to the same condition or one of the other three conditions.

All *S*s were tested individually and detailed instructions as to the nature of the task (depth-keeping) as well as information on the behavior of control-display operations were presented to them on tape. The *S*s were permitted one free trial while the instructions were being read and were allowed to ask questions prior to the start of the experiment.

During training each *S* was given 20 120-sec. trials with a 10-sec. rest between each trial within a 10-trial block. Between each 10-trial block there was a 3-min. rest period. At the completion of the training trials, a 5-min. rest period was given each *S* followed by 10 test trials. The same inter-and intratrial intervals were used for all conditions.

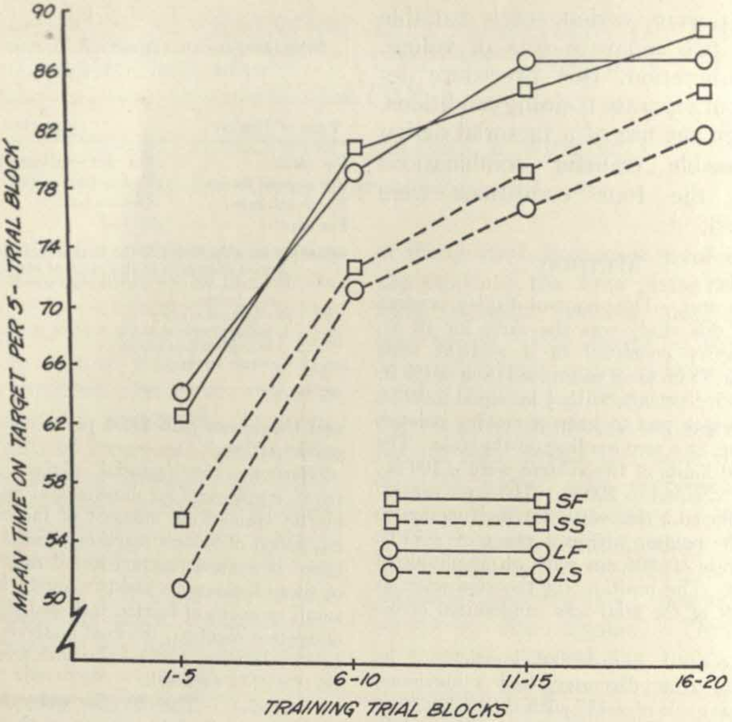


FIG. 1. Mean time on target for Cond. SF, SS, LF, and LS ($N = 20$ in each group).

The proficiency measure taken was the time on a target zone of ±10 ft. No score was recorded for the first 20 sec. of each trial to allow *S* time to bring the pointer to the ordered position. Knowledge of results was provided by informing *S* of his time on target score following each trial.

RESULTS

The dependent-variable units used in the analyses were each *S*'s mean time on target score ($\bar{X}_{\text{TOT}}$) per five-trial block. The relation between system dynamics and performance for the training data is presented graphically in Fig. 1. The general trend, clearly shown, was for better performance under Cond. SF and LF. An analysis of variance across trial blocks also was performed on the data.[3] The results showed that the main variable affecting performance was the speed of system reaction. Those groups performing the depth-keeping task with the fast speed of reaction did better than those groups operating with the slow speed ($F=4.16$, $df = 1/76$, $P < .05$). The main effect "trial blocks" also proved significant ($F=93.48$, $df=3/228$, $P<.01$), which indicated a general improvement in performance over trials.

To determine the effects of prior training on transfer, a method for evaluating these effects when *S*s are trained under different initial conditions and are performing at different levels at the time of transfer was sought. Two methods employed in studies concerned with a similar problem were: (*a*) to compare performance on the test condition for one group with performance on the initial condition for the other group (Holland & Henson, 1956; Ritchie & Michael, 1955), and (*b*) to utilize a condition of intermediate difficulty, equally unfamiliar to all groups, for the test condition (Lordahl & Archer, 1958). As neither method appeared appropriate in view of the design employed, a multivariate analysis of covariance was used to equate for differences in terminal performance levels in training prior to the imposition of test condition treatments. This analysis was performed on the first five-trial block of the test trials with the adjusting variable used being the $\bar{X}_{\mathrm{TOT}}$ scores for the last five-trial block of training trials. The test of significance of the regression of the dependent variable on the

---

[3] Bartlett's test was performed on these and all other data used in subsequent analyses with no test value being significant at the .05 level.

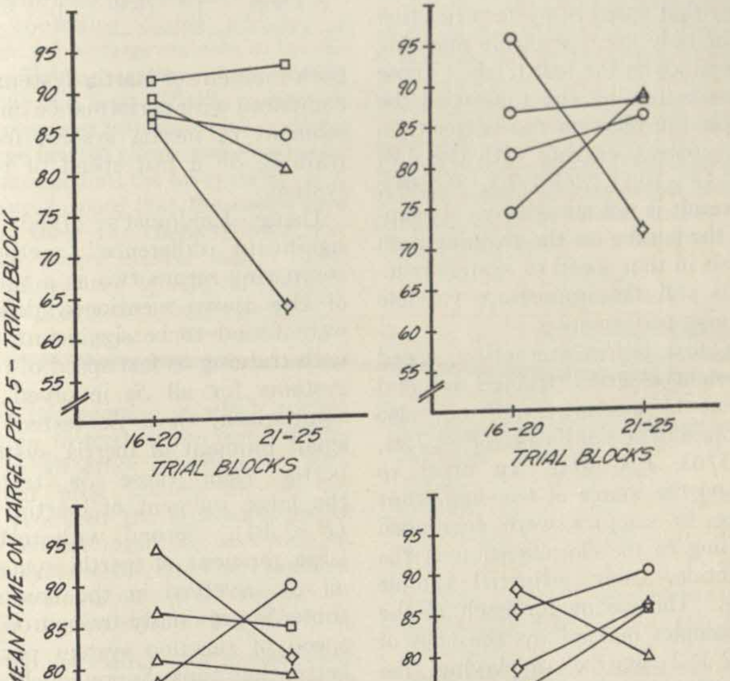


FIG. 2. Mean time on target for all transfer conditions using adjusted sample means ($N = 5$ in each group).

adjusting variable indicated a satisfactory degree of relationship ($F=46.35$, $df = 1/62$, $P < .01$). This finding supported the use of the covariance design as a means of removing part of the variance representing initial differences among the experimental units.

Using adjusted sample means the data of this analysis are plotted separately for each transfer condition in Fig. 2. The results of the analysis showed that speed of system reaction was the only main variable affecting performance on the test trials. Those groups performing the task with the fast speed of reaction did better than those groups operating with the slow speed ($F=10.11$, $df=1/63$, $P<.01$). This result is not surprising. It supports the finding on the training data analysis in that speed of system reaction is still the important variable governing performance.

The first order interaction, speed of system reaction trained on and moment of inertia tested on, also was found to be significant ($F = 7.94$, $df = 1/63$, $P < .01$). In order to focus on the source of the significant F, the 16 samples were regrouped according to the classification of the interaction, using adjusted sample means. The means for each of the four samples making up the cells of the 2 × 2 matrix displaying the interaction are presented in Table 2.

TABLE 2

MEANS OF SAMPLES REPRESENTING THE INTERACTION, SPEED OF SYSTEM REACTION TRAINED ON (SR_{TR}) × MOMENT OF INERTIA TESTED ON (MI_{TE})

Moment of Inertia Tested On	Speed of System Reaction Tested On	
	Fast	Slow
Small	87.74	81.92
Large	76.40	86.45

Inspection of Table 2 indicates that the worst performance occurred for those *S*s previously trained on a fast speed of reaction system and subsequently tested on a large moment of inertia system. The best performance was exhibited by those *S*s trained on a fast speed of reaction system and tested on a small moment of inertia system. However, training on a slow speed of reaction system resulted in better performance on both moment of inertia systems when compared with performance on a large moment of inertia system following training on a fast speed of reaction system.

Using Lindquist's (1953) "least significant difference" method for comparing means two at a time, two of the above mentioned differences were found to be significant. First, with training on fast speed of reaction systems for all *S*s involved in the comparison, those *S*s tested on the small moment of inertia system did better than those *S*s tested on the large moment of inertia system ($P < .01$). Second, with testing on large moment of inertia systems for all *S*s involved in the comparison, those *S*s previously trained on a slow speed of reaction system performed better than those *S*s previously trained on a fast speed of reaction system ($P < .05$).

DISCUSSION

The results showed clear differences in performance on a depth-keeping task favoring systems with fast speed of reaction over slow speed of reaction on both the training and test trials.[4] The differential effect of moment of inertia was not demonstrated as a main effect. It was shown, however, that moment of inertia tested on did interact with the speed of system reaction trained on.

---

[4] As no data on amplitude of stick excursion was collected, it is possible that these differences can be attributed, in part, to differences in system responsiveness. Informal observations made by *E*, however, suggest

In terms of task difficulty, these results show that dynamic systems representative of slow reacting systems are more difficult to control than those more indicative of fast reacting systems. The data also indicate that training on fast reacting systems, the "easier" task, produces performance decrements when the system controlled during transfer is characterized by large moment of inertia effects. Taken together, these results show that the most difficult system to control possesses features which combine both large inertial effects and slow speed of system reaction; the LS system. It is interesting to note that the coefficients reflecting these two characteristics produce the longer control lags. Thus, these findings tend to support the generally accepted notion that systems with long control lags are more difficult to control than those with short control lags.

As for the question of which system to train on to produce the best transfer effects to all other systems, the data show that prior training on the LS system produced the greatest positive transfer effects regardless of which of the four systems the test was made on. That is, training on the LS system resulted in no appreciable decrement on either the SF, SS, or LF systems, and did produce significantly better performance on an LS system than did prior training on any of the other systems.

These findings tend to support the general conclusion drawn by Day (1956), that transfer from a difficult task to an easy task is better than the reverse order training.

SUMMARY

Four groups of 20 *S*s were trained on a complex tracking task differing in the degree of control lag. Following training, 5 *S*s from within each group were transferred to the other systems, with one subgroup transferred to the same system. It was found that, on the basis of mean time on target scores per five-trial block, the more difficult tasks were those with the longer control lags. It was also found that training on the most difficult task, the system with the longest control lag, produced generally better transfer effects than did training on any other system.

REFERENCES

DAY, R. H. Relative task difficulty and transfer of training in skilled performance. *Psychol. Bull.*, 1956, **53**, 160–168.

HOLLAND, J. G., & HENSON, J. B. Transfer of training between quickened and unquickened tracking systems. *J. appl. Psychol.*, 1956, **40**, 362–366.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

LORDAHL, D. A., & ARCHER, E. J. Transfer effects on a rotary pursuit task as a function of first task difficulty. *J. exp. Psychol.* 1958, **56**, 421–426.

RITCHIE, M. L., & MICHAEL, A. L. Transfer between instrument and contact flight training. *J. appl. Psychol.*, 1955, **39**, 145–149.

(Received January 11, 1961)

---

that *S*s did not use full stick displacements after the first few training trials since in most cases, being naive, this behavior was unrewarding. Experienced operators can compensate for the slowness of a system reaction, but this is more likely to occur on a depth-seeking rather than a depth-keeping task.

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 376–380

EFFECTS OF EXPOSURE TIME AND DENSITY ON VISUAL SYMBOL IDENTIFICATION[1]

WARREN H. TEICHNER AND ERNEST SADLER

University of Massachusetts

In a previous study (Teichner, Reilly, & Sadler, 1961) *S*s were presented for 1 sec. with visual displays containing randomly sampled, randomly positioned groups of alphabetic letters. One group reported the number of different letters displayed (counting group); another group reported the names of the letters (naming group). The number of letters named by the naming group was then counted and compared with the number reported by the counting group. It was found that the number counted was more accurate in estimating the number displayed and that as the number of letters displayed increased, the error of the naming group increased the more rapidly. Since letter detection and discrimination were common to both groups, but the memory requirement of the naming group was much greater, these results indicate: (*a*) visual symbol identification depends on both perception and short-term memory, and (*b*) perceptual capacity is greater than short-term memory capacity.

Although both perception and short-term memory contribute to the accuracy of symbol identification, their contributions may not always be in the same direction. On the one hand, for randomly positioned letters, as the total information on the display increases, the average amount of information per fixation during scanning will also increase and this should increase the probability of letter detection since it permits more letters to be detected in fewer fixations. Up to that exposure time which allows complete scanning of the display, the perceptual process should be aided as the number of letters increases. On the other hand, findings are available (Woodrow, 1938) which indicate that there is a mutual interference of letters when they are presented in groups and this interference increases as the number of letters increases and as the distance between letters decreases.

Woodworth and Schlosberg (1954) refer the Woodrow effect to differences in the fidelity of letter images falling at or near the fixation point and those falling in peripheral vision. Although this is not unreasonable as a visual phenomenon, by the same reasoning as before, as the number of letters on the display increases, the average number of letters near the fixation point will increase and again the probability of letter detection should increase. Thus, perception should be aided rather than handicapped. But, it has been shown (Teichner et al., 1961) that the percentage of correctly identified letters decreases under these circumstances and the decrease is at least as much associated with the omission of letters as it is with incorrectly identified letters. A better explanation might relate the mutual inhibition effect to the process of storage in short-term memory (Sperling, 1960; Teichner & Myers, 1961), which would tend to offset the expected perceptual gain.

As exposure time increases, the required rate of memory storage decreases and memory should be aided. Beyond some time permitting adequate scanning of the display, perception should not be aided. For a constant exposure longer than this scanning time, increases in the number of symbols on the display should, therefore, produce a memory loss without any change in perception. The result should be an increasing loss in identification as the number of symbols (amount of information) on the display increased. Conversely, it should be possible to decrease exposure time to a point where the required storage rate is so high that recall will be a minimum. Under this condition, increases in display information should produce a negligible increased memory loss since there will be so little left to be lost. However, as noted above, the same increases will aid perception. The result of this should be an increase in identification as display information increases. Thus, it can be predicted that for very short exposure times, the accuracy of identification will increase, the greater the number of symbols to be identified, and that for longer exposure times, the reverse will be true.

---

[1] Presented to the Psychonomic Society, Chicago, September 1960. Performed under Contract N61339-588 with the United States Naval Training Device Center, Port Washington, New York.

Since the absolute number of correct identifications should increase with increasing display information, there being more symbols to be identified, a test of the predicted interaction between number of symbols and exposure time could be contaminated by a confounding of this type of effect with the interaction effect itself. It is desirable, therefore, to introduce the required perceptual and memory effects at constant information levels. This can be achieved by varying the density of the letters, i.e., by repeating the letters randomly at different parts of the display while keeping the number of different letter categories constant, and requiring that *S* report letter categories without regard to density. As letter density increases, the average proportion of the total information contained within a single fixation will increase; at the same time the number of letters will increase and their spacing will decrease, thus increasing the negative memory effect. Then, the following specific predictions can be made about the interaction of exposure time and density at any constant level of display information content: (*a*) At very short exposure times, identification will be positively related to density. (*b*) At relatively long exposure times, identification will be inversely related to density. (*c*) At some intermediate exposure time, density will have no effect. (*d*) At any density level, identification will be directly related to exposure time. The purpose of this experiment was to test these predictions at different levels of information content of the display.

METHOD

The methods used were similar to those of Teichner, Reilly, and Sadler's (1961) naming group. The display slides varied from 3 to 6 in number of letter categories and from 1 to 5 in letter density. There were 20 combinations of density and category, each of which was replicated 10 times with a different random sample of the alphabet to make up 200 slides. The time between slides was 10 sec. Two exposure times selected from 0.5, 1.0, 1.5, 2.0, 2.5, and 5.0 sec. were presented in successive halves of each session. The order of exposure times throughout the experiment was independently randomized for each of three groups of 4 *S*s each.

A 30-min. preliminary session was devoted to instructions and practice with a sample of slides not used later and with different exposure times selected from the same time range. The experimental sessions were conducted on the 3 following days. In those sessions all 200 slides were presented in random order twice, once for each of the two exposure times of the session. A different random order of slides was used for each exposure time. Each session lasted about 1 hour depending on the combination of exposure times involved. A short rest was always given between each exposure condition.

The *S*s were experimentally naive, volunteer, undergraduate students in competition for prizes of $20, $10, and $5 for the highest scores summed over all experimental sessions. No knowledge of results was given until the entire experiment was completed.

RESULTS

The data were analyzed in three ways: (*a*) in terms of the percentage of displayed categories which were correctly identified, (*b*) in terms of the percentage number of categories in *S*'s report less than the number displayed (error of omission), and (*c*) in terms of the percentage number of categories in the report greater than the number displayed (error of commission).

Errors of commission were essentially zero at a density of 1 and for greater densities never reached a value as large as 10%. Since there was also no consistent effect on these errors of exposure time, they were rejected for further analysis. Errors of omission were inversely related to percentage correct. For this reason only percentage correct data are reported.

Table 1 reports that an analysis of variance of the percentage of correct identifications found the main effects of exposure time, categories, and density all significant at the .01 level.[2] The first and second order interactions of these variables were also significant at the .01 level except those involving *S*s.

TABLE 1

ANALYSIS OF VARIANCE OF PERCENTAGE CORRECT IDENTIFICATIONS

Source	*df*	*MS*	*F*
Between *S*s	11		
Within *S*s	1428		
Exposure Time (T)	5	4.510	50.11*
Categories (C)	3	3.880	485.00*
Density (D)	4	.050	6.25*
T × C	15	.050	8.33*
T × D	20	.028	4.67*
C × D	12	.028	3.50*
T × C × D	60	.005	1.67*
S × T	55	.090	
S × C	33	.008	
S × D	44	.008	
S × T × C	165	.006	
S × T × D	220	.006	
S × C × D	132	.008	
S × C × D × T	660	.003	
Total	1439		

$^*P < .01$.

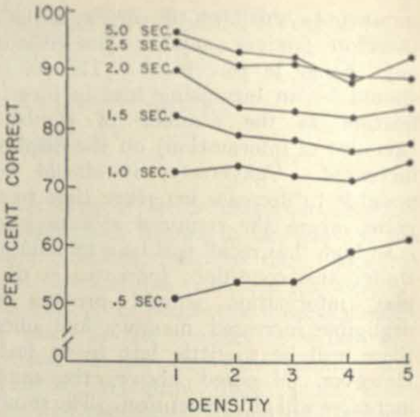


FIG. 1. Percentage correct as a function of density with exposure time as a parameter.

Figure 1, which presents the effects of density and exposure time pooled over categories, shows that accuracy increased with increasing exposure time up to at least 2.5 sec. The small difference between the 2.5-sec. and 5.0-sec. times suggests a limit somewhere between these two values. When these data are replotted on a time base, percentage correct is a negatively accelerated increasing function of exposure time with an asymptote of approximately 2.5 sec.

Figure 1 also reveals the nature of the interaction between exposure time and density. At the shortest exposure time percentage correct was essen-

---

[2] Complete data by *S*s for all conditions have been deposited with the American Documentation Institute. Order Document No. 7035 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

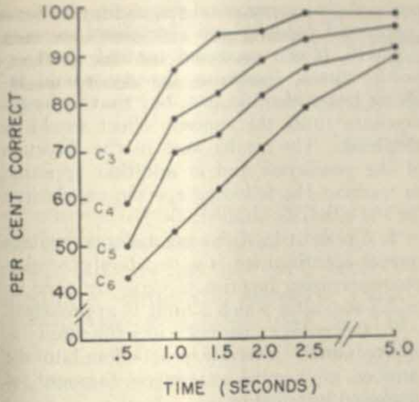


FIG. 2. Percentage correct as a function of exposure time with categories as a parameter.

tially directly proportional to density. At the 1.0-sec. exposure no general trend is apparent. At exposure times greater than 1.0 sec., there is a downward trend indicating an inverse relationship. The data suggest that the slopes of these downward trends tend to increase with increasing exposure time. The slopes are small in all cases.

Figure 2, which presents percentage correct as a function of exposure time, pooled over density, with number of categories as the parameter, shows that percentage correct was inversely related to number of categories and was a negatively accelerated increasing function of time. The asymptote of these time functions appears to decrease with increasing time; except for the six-category case, increases beyond 2.5 sec. are small.

Figure 3 presents percentage correct as a function of categories with exposure time as the parameter. This figure indicates that accuracy was in general inversely proportional to number of categories; the effect of exposure time was to reduce the slope of the function.

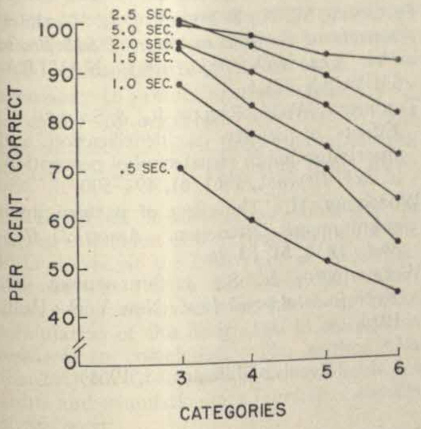


FIG. 3. Percentage correct as a function of number of categories with exposure time as a parameter.

DISCUSSION

The results appear to be as predicted. At the 0.5-sec. exposure time identification increased as density increased. This was predicted on the basis of a perceptual gain due to the increased proportion of the total information contained within a fixation and the consequent increase in the probability of letter detection. At this short exposure time, short-term memory was presumably approaching a lower limit; thus very little further loss would be possible as the effect of increasing interference. At the 1.0-sec. exposure time that effect was sufficiently important to nullify the perceptual gain and at longer exposure times, it overrode it. The data suggest that by 2.5 sec., exposure time was generally long enough to provide for adequate scanning so that further improvements in perception were small. On the other hand, the memory interference due to density was still present at the longest exposure time.

The results obtained are in good agreement with correlated points of the previous experiment (Teichner et al., 1961) which also found no density effect for the 1-sec. exposure time and which found the same kind of relationship between identification and the number of letter categories.

Density, as we are using the term, is a form of redundancy and in a noisy channel, redundancy would be expected to increase information transmission (Shannon & Weaver, 1949). On the other hand, in an essentially noise-free channel, as is the case here, there are psychological studies which show that redundancy in the form of density has a decrement-producing effect (Christner, Schutz, & Ray, 1959; Green & Anderson, 1956). Our study, then, is relevant to the explanation of these results and suggests that where redundancy increases information transmission there is a relatively large gain in the rate of transmission along a perceptual channel and otherwise there is a more important interference with short-term memory. This approach also suggests, for the same reasons, that the memory process is largely responsible for the loss in information transmitted which is commonly found with increases in stimulus information. Thus, the present study extends previous ones which have shown that *S* perceives more than he can recall (Sperling, 1960; Teichner & Myers, 1961; Teichner et al., 1961) by demonstrating the nature of the interaction between the two in their effects on identification.

Finally, regarding density, although redundancy is considered a means of increasing transmission along a noisy channel, so far as we know, it has not been suggested that redundancy might be a means of increasing transmission along an already overloaded but noise-free channel. Our results suggest that for the human channel and the special case of a short exposure time, this may be reasonable, at least for the form of redundancy used here.

SUMMARY

Twelve *S*s were presented with displays of alphabetic letters which varied in number of different letter categories, letter density, and exposure time. On the basis of a previously-made experimental distinction between perceptual capacity and short-term memory capacity, it was predicted that at short exposure times, increasing density would increase letter identification, but that at longer exposure times the opposite effect would be obtained. The results were in the direction of the prediction and in addition appeared to warrant the following specific conclusions for the situation studied:

1. Up to at least five symbols, percentage correct identification is a negatively accelerated increasing function of exposure time up to 2.5 sec. after which a limit is approached.
2. Percentage correct identification is approximately inversely proportional to the number of symbol categories (amount of displayed information).
3. The percentage of omitted symbols is a negatively accelerated decreasing function of exposure time with a limit of approximately 2.5 sec.

REFERENCES

CHRISTNER, C. A., SCHUTZ, R. G., & RAY, H. W. Some factors affecting visual search time for symbols on a large display. Paper read at American Psychological Association, Cincinnati, Ohio, 1959.

GREEN, B. F., & ANDERSON, L. K. Color coding in a visual search task. *J. exp. Psychol.*, 1956, **51**, 19–24.

SHANNON, C. E., & WEAVER, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74** (11, Whole No. 498).

TEICHNER, W. H., & MYERS, J. L. Training aspects of decision making. *USN Train. Dev. Cen. tech. Rep.*, 1961, NAVTRADEVCEN 588-1.

TEICHNER, W. H., REILLY, R., & SADLER, E. Effects of density on identification and discrimination in visual symbol perception. *J. exp. Psychol.*, 1961, **61**, 494–500.

WOODROW, H. The effect of pattern upon simultaneous letter-span. *Amer. J. Psychol.*, 1938, **51**, 83–96.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received January 13, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 381–386

SENSORY INTEGRATION WITH AND WITHOUT REINFORCEMENT[1]

VERNON O. TYLER, JR.[2]

University of Nebraska

Sensory integration or S-S learning has been a notion of interest for some time, but it has not yet received adequate empirical support (MacCorquodale & Meehl, 1954, p. 213; Spence, 1950). Osgood (1957) cites evidence for the concept, but is not fully satisfied with these findings. The purpose of the present study was to test the notion once again, but with a more rigorous methodology.

Osgood postulates three kinds of learning: S-R, R-R, and S-S. In his theory, S-R learning is a function of a number of factors, with reinforcement an important one; S-S and R-R learning are said to be relatively independent of reinforcement with contiguity postulated as sufficient for the establishment of associations.

In this discussion, S will be used to represent within-the-organism stimulus correlates or traces resulting from the presentation of a stimulus and R to represent within-the-organism precursors of responses or traces of these precursors.

To demonstrate S-S learning, it is necessary to present stimuli contiguously in order to ascertain whether or not associations between their corresponding Ss are formed. At the same time, however, it is necessary to make certain that no S-R or R-R associations are formed or strengthened. This can be done by preventing responses from occurring or by making certain that overlearning has previously taken place so that any existing S-R associations cannot be strengthened.

Following the contiguous presentation of a number of stimuli, the presence of sensory integrations among their stimulus correlates may be ascertained by presenting some of the stimuli and observing whether or not there is evidence that the correlates of the remaining stimuli are activated in some way. At the present time, the only tangible evidence that these remaining stimulus correlates have in fact been activated is by the occurrence of some signifying or indexing responses linked to these correlates.

Sensory integration then will be defined here as the associations which develop among the internal stimulus correlates of a group of stimuli (or parts of a stimulus complex) which are presented together such that a part of the group, in a later presentation, will elicit a set of indexing responses which have occurred previously only to the total original group of stimuli. This definition is essentially the same as Osgood's, although it is more explicit.

This procedure poses several methodological problems: (*a*) Since the indexing responses must be suppressed during the training phase of this procedure, it is difficult to ascertain that the stimuli whose correlates are to be integrated do in fact impinge upon the receptors in an effective manner. For example, in the visual modality, *E* must be sure that *S* actually "sees" the stimuli. If, in a test for sensory integration, positive results are obtained, it would appear obvious that the stimuli must have impinged upon the receptor involved. Negative results, however, would leave this question unanswered. (*b*) Another problem is that of controlling the indexing responses during training. This is necessary in order to exclude S-R and R-R interpretations of positive findings. This would mean making certain that no new associations between Ss and Rs (S-R learning) or between Rs (R-R learning) involving indexing responses were developed during S-S training. Solution to this problem could involve using indexing responses each one of which has an overlearned association with one and only one stimulus and little or no association with any other stimuli, and by having *S*s make competing responses to the stimuli during the training. (*c*) Another problem is that of developing a sensitive means of measuring the indexing response during the test phase of the experimental procedure. In this experiment the tachistoscope was used because it appeared well suited for controlling the amount of "information" received by *S* over a considerable range and in sufficiently fine gradations to make a test sensitive even to minute variations in the strength of S-S associations.

---

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Nebraska. The author thanks Katherine E. Baker, dissertation supervisor, for her help in the formulation of this design and in seeing this research to completion. The author also thanks John C. Loehlin for his valuable insights and stimulation on a number of aspects of this work.

[2] Now at Fort Worden Diagnostic and Treatment Center, Port Townsend, Washington.

It is the failure to control responding which leaves most of the previous work in this area inconclusive. In Solomon and Postman's (1952) study, responding was clearly present with *S* pronouncing nonsense words. The same is true of Cohn's (1954) work. In Gibson's studies (Gibson, Walk, Pick, & Tighe, 1956) in which animals had prolonged exposure to visual forms, responses while not differentially reinforced were certainly not rigidly controlled. Nor does Postman and Conger's (1954) study fully control the recurrence of a response—certainly not with laboratory precision.

The primary hypothesis tested in this study was that tachistoscopic thresholds are lower for recognition of nonsense words to which *S*s have been previously exposed (with indexing responses absent during training) than for nonsense words which are new to them (control words). It was also hypothesized that thresholds are independent of reinforcement during training, at which time S-S integrations are presumably being formed.

PROCEDURE

Word lists and apparatus.—A large pool of five-symbol, nonsense "words" consisting of four capital letters and one Arabic numeral was constructed. The letters were randomly assigned to the words so that no letter was repeated until the alphabet was exhausted and no letter was repeated in any one word. The third symbol of each word was a randomly assigned digit from 1 to 9. These nonsense words were typed on a paper tape which was fastened to rollers in a Gerbrands mirror-type tachistoscope.

The tachistoscope was employed to check the visibility of the words. The ascending method of limits with .01-sec. steps was used to determine the duration of exposure necessary for *S*s to identify correctly four out of the five symbols (letters or digits) making up each nonsense word, ignoring any additional symbols which might be erroneously guessed to be present. This lower-than-usual criterion was used because it permitted greater sensitivity in detecting small differences in visibility of words. The illumination in the tachistoscope as measured by a Macbeth Illuminometer was 1.3 apparent ft-c at the eyepiece for both the "blank" field and the stimulus exposure field. The tachistoscope was situated in a small experimental room illuminated by a 100-w. red light bulb to minimize loss of dark adaptation when *S* looked away from the eyepiece.

From preliminary experimentation, a number of words were selected for which approximately equal thresholds were obtained. From this list, two pairs of words were selected so that the members of each pair were as nearly equal as possible in means and ranges of the thresholds they yielded. The four words of these two pairs were made into a final list ordered on the basis of obtained thresholds beginning with the word giving the lowest average threshold and ending with the word giving the highest average threshold. This order was chosen with the object of obtaining fairly uniform thresholds by balancing out the tendency for thresholds to decrease with ordinal position in the list (Howes & Solomon, 1951) and individual word differences in threshold based on their inherent differ-

TABLE 1

LIST USED FOR OBTAINING THRESHOLDS AND USE OF EACH WORD FOR TWO GROUPS OF *S*S

Words	Group I	Group II
1. NA8XV	Practice	Practice
2. GT7BZ	Practice	Practice
3. LQ2UZ	Practice	Practice
4. IA9SJ	Practice	Practice
5. WY9OQ	Experimental	Control
6. GW6QB	Control	Experimental
7. YA5TI	Experimental	Control
8. LX7OJ	Control	Experimental

ences in visibility. In addition, four high-threshold practice words were placed at the head of this list, which is presented in Table 1.

This list was presented to all *S*s in the main experiment. Words 5 and 6 comprised one pair of equated words and 7 and 8 another pair. Words 5 and 7 were the experimental words for Group I with Words 6 and 8 serving as controls. For Group II, Words 6 and 8 were experimental words and 5 and 7 served as control words. The use of this counterbalanced design precluded attributing any positive results to differences in word difficulty or serial position.

Preliminary training involved presenting two of these words together with other "filler" words. For this training, two sets of 60 3 × 5 in. training cards were prepared. A nonsense word was typed in the center of each card. The 60 cards of Set I (presented to *S*s of Group I) consisted of 25 cards with Word 5 (from the list of Table 1) typed on them, 25 cards with Word 7, and 10 cards, each with a different word, which served as the fillers. Each filler card had 1 of 10 nonsense words randomly drawn from the original pool. The 60 cards of Set II (presented to Group II) consisted of 25 with Word 6, 25 with Word 8, and 10 with the same filler words as Set I. The cards of each set were thoroughly shuffled before every administration.

Subjects.—Twenty-eight *S*s participated in the main experiment, 16 in Group N (neutral) (3 men and 13 women), 12 in Group R (reinforcement) (8 men and 4 women). Both of these groups were in turn divided into halves, Groups I and II, in order to counterbalance experimental and control words as described above. Most of the *S*s were high school students who volunteered for the experiment while a few others were volunteers with whom *E* was acquainted. Age varied from 13 to 22 yr., with most of the *S*s between 15 and 19 yr.

Procedure.—The general procedure was to present one of the sets of cards to an *S*, a card at a time, and then take him to another room and obtain his tachistoscopic thresholds. The *S*s from both Groups N and R were randomly assigned to Group I or II and presented the cards of the corresponding set, I or II. The *S* was taken to a small, well-illuminated experimental room with a table and two chairs. A wooden screen on the table separated *S* from *E* so that he could not be observed "recording data." The instructions, paraphrased, were as follows:

> In this research, we are interested in how clearly things appear to people. This is important in such things as making instrument panels for aircraft and automobiles. Take this stylus (*E* hands stylus to *S*), point to each letter here in turn, and say whether it is "round" or "straight." That is, if a letter is made up of round lines or a combination of round and straight lines, call it "round." If it is made up of only straight lines, call it "straight." Just move the stylus along saying "round" or "straight." Relax and go at your own speed.

It should be mentioned here that the type of response required of *S* was intended to to prevent his subvocalizing the names of the letters. Two pilot studies indicated that this procedure was effective. In one of these studies, for each of the letters in a series of nonsense words, each *S* was deliberately instructed to make the indexing response in a covert manner and then to make overtly the presumably competing response. The *S* was first instructed to "think" the names of each of the letters in the nonsense words and then say whether each letter was "round" or "straight." After he had gone through this procedure, he was asked to say *only* whether the letters were round or straight, going through the word. With 16 *S*s the mean time (7.82 sec.) required for thinking names of letters *and* making competing responses was significantly greater than that required for making competing responses alone (4.42 sec.), using a t test for correlated measures ($P < .005$). These data indicate that the competing responses effectively prevented even the covert exercise of indexing responses.

In the second pilot study, it was also shown that indexing responses were not covertly rehearsed while *S*s were making competing responses. This was done by demonstrating that it takes *S* no longer to make responses such as counting or describing letters than it does to make the same type of responses to completely new, unfamiliar stimuli for which no old, overlearned response labels are available. If it takes as much time to make a given type of competing response to both familiar and unfamiliar stimuli, then it is doubtful that an *S* has "smuggled in" any implicit indexing response to the familiar symbols during training. This procedure was utilized with 8 *S*s and a comparison was made of the time required to describe letters as either round or straight and to describe nonsense figures which were equivalent to the letters with respect to the number of curved and straight elements. The mean time required for describing the two kinds of stimuli did not differ significantly (5.06 and 5.09 sec., respectively, $P > .90$), using a *t* test for correlated measures. These results were taken as further evidence that *S*s do not subvocalize names of letters while describing them under these training conditions.

Returning to the experimental procedure, it should be said that *S*s in Group R were given an approving word or phrase immediately after they completed their performance of the task of naming round and straight symbols on each card. These phrases, randomly used, were such expressions as "Good," "Excellent," "Fine," and "Well done."

After presentation of the cards, *S* was immediately taken to another experimental room and seated in front of the tachistoscope. He was told (paraphrased here):

> Here we are studying the effects of lighting on the visibility of letters and numbers. Now look into this eyepiece. Do you see the two horizontal lines? I am going to flash a word between them for a brief interval and I want you to guess how it is spelled. It is not an ordinary word but a nonsense word—just a random arrangement of letters and numbers. The first time I flash the word, you probably won't be able to see much—nothing more than a flash of light. However, I will keep flashing the word for longer and longer intervals until it will be easier to recognize the letters and numbers. Start guessing as early as possible, even though you're not sure.

Before each exposure, *E* said "Ready?"

Thresholds were then obtained on all eight words in the list, using the ascending method of limits. The tachistoscope was first set to flash a word for a duration of .01 sec. The second exposure was for .02 sec., and so on in .01-sec. increments. For the four practice words, this procedure was continued until *S* "spelled" out all five symbols of the word perfectly or until an exposure of .80 sec. was reached. Only 2 *S*s failed to spell a word correctly before this limit of .80 sec. This use of the 5/5 symbols criterion for the first four words was to give *S*s more practice and to emphasize accuracy before they reached the critical words.

Thresholds for the remaining four words (Words 5 to 8) were obtained with the same procedure except that the criterion was 4/5 symbols correctly guessed in the proper order, but with extraneous intervening symbols permitted.

RESULTS AND DISCUSSION

This procedure yielded four threshold measures for each *S*, two for control words of no prior exposures and two for experimental words with 25 prior exposures each. The scores for the two control words for each *S* were arithmetically averaged and, likewise, scores for the two experimental words for each *S*. Thus each *S* contributed two scores (two averages) to the statistical treatment. Means and *SD*s for each of the four experimental conditions are presented in Table 2.

It is apparent from these data that (*a*) both Groups N and R had lower thresholds for experimental words than for control words; (*b*) Group R had lower thresholds than Group N for both experimental and control

TABLE 2

MEAN THRESHOLDS (SEC.) FOR FOUR EXPERIMENTAL CONDITIONS

Group	Control Words (Frequency =0)		Experimental Words (Frequency =25)	
	Mean	*SD*	Mean	*SD*
Neutral	.1413	.036	.1194	.041
Reinforcement	.1221	.036	.1054	.030

Note.—Original thresholds were in .01 sec., *S* means in .001 sec., and Group means in .0001 sec.

TABLE 3

Analysis of Variance of Thresholds

Source	*df*	*MS*	*F*
Between *S*s	27		
Reinforcement (R)	1	.00376202	1.739
Error (b)	26	.00216382	
Within *S*s	28		
Frequency (F)	1	.00540175	8.093*
R × F	1	.00009304	<1
Error (w)	26	.0006675	

* $P < .01$.

words; and (*c*) the slopes of the curves for the two groups, if plotted graphically, would be essentially the same.

To evaluate the data statistically, they were subjected to the analysis of variance presented in Table 3. The predicted Frequency effect (experimental vs. control words) is highly significant. This indicates that exposure to the nonsense words (experimental words) reduced tachistoscopic thresholds for their recognition significantly below thresholds for words *S*s had not seen before (control words). Thus the primary hypothesis of this study postulating a sensory integration type of learning is supported. In addition, the direction of the difference of the Frequency effect in these data supports the prediction that tachistoscopic thresholds are a decreasing function of the frequency of exposure; or, in the general terms of the theory, they support the hypothesis that the strength of sensory integration is an increasing function of the frequency of contiguous presentation of parts of the stimulus complex.

Although the Reinforcement effect is not statistically significant, there is a trend in the direction suggesting that reinforcement was serving to raise the general level of *motivation* of reinforced *S*s, resulting in lower thresholds on all words (experimental *and* control).

The effect of reinforcement on learning, as opposed to performance, is indicated by the Frequency × Reinforcement interaction. It is clear there is no such interaction effect. This indicates that while Group R performed somewhat better (i.e., had lower thresholds) than Group N, the superiority of Group R to Group N was about the same for both experimental *and* control words. Since only the experimental words in Group R were reinforced, the control words not being given during training when the reinforcement occurred, and since both experimental and control words in this group show about the same margins of superiority over the corresponding experimental and control words of Group N, it was concluded that reinforcement did not strengthen sensory integration. The fact that no significant interaction effect occurred supports the second hypothesis of this study, namely Osgood's assumption that reinforcement has no effect on the *strength* of sensory integration and probably is not required even for the *formation* of sensory integration.

In general, the results of this study appear to provide ample evidence for the validity of the notion of sensory integration or S-S learning. As far as can be discovered from the literature, they represent the first data that unequivocally show that sensory integration can be demonstrated experimentally under appropriately and adequately controlled laboratory conditions.

These findings throw new light on the question of whether a response is necessary for learning. It seems reasonable to conclude that no response is necessary for at least one kind of learning. This conflicts with Hullian theory which utilizes only S-R learning and supports Tolman in his espousal of S-S learning.

Moreover, these findings raise once again a point that Tolman has stressed for some time. Perhaps many experimental situations which have been taken as instances of pure S-R learning may best be interpreted as S-S learning subtly contributing to responding. Now that greater attention has been given to controlling responding, typical S-R situations can be restudied to see what aspects of performance changes might take place when responding is controlled during preliminary training. This does not say that S-R conceptualizations are going to be eliminated by this approach, but rather it points out the need to delimit the area to which S-R notions alone are sufficiently explanatory. If different kinds of learning do follow different laws, certainly this calls for mapping their particular spheres of functioning in the organism.

SUMMARY

The purpose of this study was to demonstrate S-S learning with particular care to control responding and thereby exclude S-R and R-R interpretations of the results. In the training phase of the experiment, *S*s were asked to point to the letters of nonsense words and describe them in a way that made certain that the names of the letters were not uttered vocally or subvocally, thus ruling out S-R and R-R interpretations. In the test phase, tachistoscopic duration thresholds were obtained for the nonsense words and equated control words.

Thresholds were significantly lower for previously exposed words than for control words, which was interpreted as evidence for S-S learning. One group of *S*s was praised during the training phase and did no better on the tachistoscope than an unpraised group. This was interpreted as indicating that S-S learning is independent of reinforcement.

REFERENCES

COHN, B. The effects of frequency and consequence as revealed by several different measures of verbal learning. *J. abnorm. soc. Psychol.*, 1954, **49**, 290–297.

GIBSON, E. J., WALK, R. D., PICK, H. L., & TIGHE, T. J. Effects of prolonged exposure to visual patterns. *J. comp. physiol. Psychol.*, 1956, **49**, 239–242.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.

MACCORQUODALE, K., & MEEHL, P. E. Edward C. Tolman. In W. K. Estes, S. Koch, K. MacCorquodale, C. G. Mueller, Jr., W. N. Schoenfeld, & W. S. Verplanck (Eds.), *Modern learning theory*. New York: Appleton-Century-Crofts, 1954. Pp. 177–266.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 348–424.

POSTMAN, L., & CONGER, B. Frequency of stimulation and frequency of response as factors in recognition. *Science*, 1954, **119**, 671–673.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.

SPENCE, K. W. Cognitive versus S-R theories of learning. *Psychol. Rev.*, 1950, **57**, 159–172.

(Received January 13, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 387–390

A PSYCHOPHYSICAL STUDY OF HUNGER IN THE RAT[1]

ROBERT C. BOLLES

Hollins College

There is a growing tendency to attribute motivational phenomena to the *stimulus* consequences of deprivation (Estes, 1958). One implication of this viewpoint is that it ought to be possible to do psychophysical experiments with the alleged deprivation-produced stimuli. It is a source of some little embarrassment therefore that drive discrimination and especially drive-intensity discrimination have been typically so difficult to demonstrate (Bloomberg & Webb, 1949; Jenkins & Hanratty, 1949). The purpose of the present study was to demonstrate that the difficulties encountered in previous studies are mainly methodological, and that an experimental psychophysics of hunger is feasible. Specifically, animals were trained to discriminate between large and small prefeedings; the discrimination was made increasingly difficult so that "just noticeably different" prefeeding values could be determined; and finally, the obtained jnd's were added to derive a psychophysical scale of hunger.

METHOD

Subjects.—The *Ss* were 54 male albino rats, approximately 100 days old at the start of training. Fifty *Ss* survived the 3 mo. of training and testing.

Apparatus.—Each *S* was confined for the course of its training and testing in an individual unit which consisted of a home cage, two alleys leading off the home cage in opposite directions, and goal boxes at the ends of the alleys. The home cage was 10 × 10 in. and 4 in. high, with hardware-cloth over sawdust flooring, and sliding doors closing off the alleys. The alleys were 4 in. wide, 4 in. high, and 4 ft. long and they terminated in goal boxes that were 4 in. wide, 4 in. high, and 10 in. long. Nonretrace doors separated the goal boxes from the alleys. The whole unit was made of unpainted wood, except for one side which was covered with hardware-cloth. Thus, *S* was provided with good spatial-orientation cues, since the two goals were well-separated in space, and since the two alleys presented extramaze cues on opposite sides of *S* as it ran. Leeper (1935) and Bolles and Petrinovich (1954) have indicated that good spatial-orientation cues can greatly facilitate drive discrimination. Twelve of these units were stacked together so that as many as 12 *Ss* could be tested at a time.

Training.—Throughout the study hunger and also, presumably, the to-be-discriminated stimulus consequences of hunger were controlled by prefeeding *S* a certain proportion of its average daily consumption. This procedure has the advantage over the more usual procedure involving deprivation for a certain number of hours since it provides for more rapid onset of the relevant stimuli. Prior to the discrimination training *Ss* had had a few days of 23-hr. deprivation experience, and had been given 2 days in which to explore the apparatus, 1 day with access to ad lib. food on one side, and 1 day with access to ad lib. food on the other. During training *S* was prefed in its individual home cage, where it had been confined for approximately 23 hr. When the prefeeding ration had been eaten (within 20 min., ordinarily) the food dish was removed from the home cage and *S* was detained there another 5 min. Then the sliding doors leading to the two alleys were withdrawn and *S* ran its one trial for the day. Food to finish the daily meal was in one goal box on days when the prefeeding ration was large, resulting in "low hunger," and in the other goal box on days when the prefeeding ration was small, resulting in "high hunger."

Large and small prefeedings differed between *Ss* to some extent because prefeedings

---

[1] The present study was made possible by the continuing support of the National Science Foundation. Preliminary findings were reported in an unpublished bachelor's thesis, by Henry Calam, Princeton University, and at the 1959 meeting of the Midwestern Psychological Association.

were only measured to the nearest .5 gm., and there was some variability among *S*s in total consumption, but the large prefeeding amounted to about seven-eighths, and the small prefeeding to about one-eighth, of *S*'s average total daily intake. Thus, *S* was required to discriminate between high and low hunger where the difference represented about three-fourths of its daily food intake. High and low conditions and left and right sides were counterbalanced across *S*s. Except as noted below, *S*s were run under high and low hunger on alternate days. The "correction method" was used, so *S* could retrace after an error. An error was counted only if *S* ran far enough down the wrong alley to be able to see into the empty goal box, i.e., as far as a low barrier placed 12 in. from the end of the goal box. After entering the correct goal box, *S* was confined there until it had finished eating. The time of confinement stabilized during the course of training as *S*'s goal box consumption stabilized at some value appropriate for the size of its prefeeding. The goal box food dish always contained an excess of available food, and food was weighed regularly before and after eating so *S*'s gradually increasing consumption could be taken into account in preparing constant proportion prefeedings. A wet mash consisting of two parts water to one part ground Purina lab chow was used. A water bottle was always available in the home cage but *S*s drank very little, rarely as much as 5 cc per day, because of the high water content of the mash. After the one daily trial, the doors were opened and *S* was allowed to run back to its home cage where it was confined without food until the next day. Discrimination training was continued until *S* met a 9 out of 10 criterion of learning.

Testing.—When the discrimination was established *S* was run on successive blocks of 10 trials where the difference between large and small prefeedings was progressively reduced. Thus, the discrimination was made progressively more difficult until, finally, as the prefeedings were made the same size, the task became impossible. Ideally, the discrimination performance would reflect the difficulty of the task, dropping from an initial level of 90% correct to a chance level, thereby yielding a psychometric function. The prefeeding differential at which performance was 75% could then be defined as a prefeeding jnd, a discriminability threshold.

Six different groups of *S*s were run to determine discriminability thresholds at six different points on the prefeeding continuum. Thus, for one group large and small prefeedings were brought progressively toward one-half the daily consumption in order to determine the threshold at the middle of the prefeeding scale. Two groups of *S*s were run to obtain threshold determinations at the extreme ends of the scale, and three groups were run to fill in intermediate points. Except for the variation in size of prefeeding, the same procedures were followed that had been used during the initial training.

Several precautions were taken to insure that *S*'s correct responding was contingent upon discrimination of its hunger, or at least upon the size of its prefeeding, rather than upon some cue inadvertently provided by *E* or the test situation. For example, the daily alternation of correct sides was frequently interrupted. Although *S*s in a group were run within a few minutes of each other, the order of running them was constantly changed. As a consequence *E* typically did not know until *S* had made its daily run which side was correct.

RESULTS

The median *S* (of the total 54) met the 9 out of 10 criterion of learning in 43 trials, and 33 *S*s learned the discrimination within 60 trials. These 33 *S*s then served under the six different sets of conditions used to determine discriminability thresholds at six different points on the prefeeding continuum. Either 5 or 6 *S*s were tested under each set of conditions. Figure 1 gives the results for one group of *S*s for which large and small prefeedings were brought progressively closer to 50% of the mean daily consumption. Each point on Fig. 1 represents the performance on a block of 10 trials of an individual *S* under a particular prefeeding differential. The figure is drawn as though all *S*s were tested under the same conditions of differential prefeeding. Because prefeeding rations were only measured to the nearest .5 gm., and because there was some individual variation in total consumption, it was not possible to keep prefeeding differentials exactly the same for all *S*s; but they were kept within about 5%, and the median values are plotted in Fig. 1. The median percentage correct for the group at a given median prefeeding differential is also shown. The ordinate is plotted in normal deviate units so that the psychometric function could be drawn linearly to fit the group medians. The obtained discriminability threshold, the difference between "just noticeably different" values, amounted to 28% of the total daily food consumption.

For this group of *S*s high hunger had been reduced and low hunger had been increased symmetrically, so that, at threshold, *S* was just able to discriminate between a large prefeeding, amounting to 64% of its total daily food consumption, and a small prefeeding, amounting to 36% of its total daily food consumption. When the large and small prefeeding rations were equated, and made equal to 50% of the daily total, then the median performance for the group fell to 45% correct.

Discriminability thresholds were obtained at five other locations on the prefeeding (hunger) continuum. They varied in size from 40% (that is, the difference between 90% and 50%) toward the low hunger end of the continuum to 20% (the difference between 20% and 0%) at the high hunger end. These findings are summarized in Fig. 2. Taking the discriminability threshold, or jnd, as a unit of hunger intensity, a psychophysical scale of hunger may be derived in the classical Fechnerian manner: by adding up the empirically obtained jnd's and plotting them against their corresponding physical magnitudes (prefeeding rations). Figure 2 presents a smooth curve fitted to the present threshold determinations.

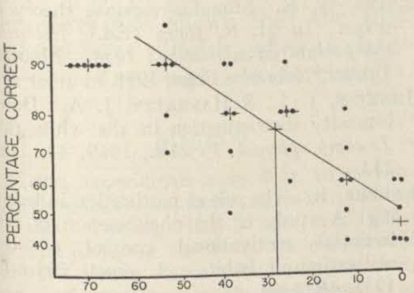


FIG. 1. Individual and group discrimination performance as a function of the size of differential prefeeding.

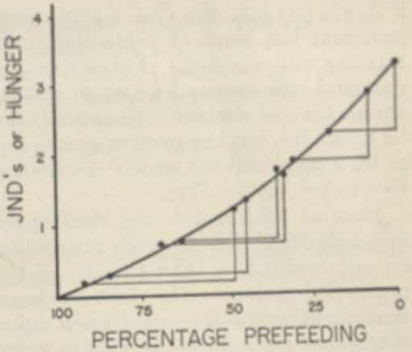


FIG. 2. Discriminability of hunger as a function of the percentage of daily food given in prefeeding.

DISCUSSION

Drive-intensity discrimination was obtained with far fewer trials in the present study than in the previous studies by Jenkins and Hanratty (1949) and Bloomberg and Webb (1949). This faster learning may probably be attributed to certain procedural advantages: (*a*) There were good extramaze cues for spatial orientation; (*b*) the onset and termination of the stimuli to be discriminated occurred at the choice point and in the goal box (Bolles & Petrinovich, 1954); (*c*) the onset and termination of the stimuli to be discriminated occurred quickly (achieved here by using a prefeeding procedure and feeding a wet mash that could be quickly consumed); (*d*) *S* lived at the choice point. (Preliminary investigation had shown that if *S* was put into the choice point compartment just prior to prefeeding, then learning was retarded.) Further methodological refinements leading to smaller thresholds and simpler experiments may be hoped for, but the present procedures at least make drive-intensity discrimination technically feasible.

None of the control procedures gave any indication that *S*s were responding on any basis other than the size of the prefeeding ration. The best evidence that performance was contingent upon differential prefeeding was that in every case performance broke down on the impossible task. A few *S*s were put back on an easy discrimination task at this point. Some retained the discrimination, others had to be retrained. In short, there seems little doubt but that *S*s were responding discriminatively to the differential prefeeding.

Some doubt might arise about whether *S*s were actually hungry rather than thirsty, since they were receiving most of their daily water at the same time they were consuming their daily food. Upon the completion of testing, some *S*s were tested with water in one goal box and dry Purina pellets in the other. It was invariably found that water was refused but the dry food was accepted. Hence, it does not seem reasonable to suppose that *S*s were primarily thirsty. It should be recognized however that the particular results presented here in Fig. 2 might have been somewhat different if the food used had not been so wet.

The psychophysical function of hunger given in Fig. 2 is concave upward, in contrast with previous attempts to quantify hunger where curves convex upwards have generally been reported. The discrepancy may depend upon the fact that in the present study percentage prefeeding rather than the more usual hours of deprivation was the independent variable. Only further experimentation can determine whether these two experimental procedures consistently produce similarly discrepant quantitative effects. Similarly, further research will be necessary to establish whether the construct scaled and labeled "hunger" in Fig. 2 will have empirical correlates such that we will want to call it hunger. Another line of attack upon the problem of defining hunger might well start with a search for the discriminative stimuli in the present situation. Are they in the stomach, or the head, or somewhere else? Either approach to the problem might permit the much abused and ill-defined word "hunger" to be pinned down quantitatively in terms of the stimulus consequences of deprivation.

SUMMARY

Rats were trained to respond differentially on the basis of small and large prefeedings. When the discrimination had become established (the median trials to criterion was 43) the task was made progressively more difficult by reducing the difference between small and large prefeedings. The point at which performance fell to 75% correct was defined as a discriminability threshold, and such thresholds obtained at different points on the prefeeding continuum were used to construct a psychophysical "scale" of hunger.

REFERENCES

BLOOMBERG, R., & WEBB, W. B. Various degrees within a single drive as cues for spatial response learning in the white rat. *J. exp. Psychol.*, 1949, **37**, 628–636.

BOLLES, R., & PETRINOVICH, L. A technique for obtaining rapid drive discrimination in the rat. *J. comp. physiol. Psychol.*, 1954, **47**, 378–380.

ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958.

JENKINS, J. J., & HANRATTY, J. A. Drive intensity discrimination in the white rat. *J. comp. physiol. Psychol.*, 1949, **42**, 228–232.

LEEPER, R. The role of motivation in learning: A study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.

(Received January 14, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 391–395

KEEPING TRACK OF VARIABLES THAT HAVE FEW OR MANY STATES

DOUWE B. YNTEMA AND GAYLE E. MUESER

Lincoln Laboratory,[1] *Massachusetts Institute of Technology*

This article describes further experiments on keeping track of several variables. The *S* is read a series of messages, each of which tells him that one of the variables is, and until further notice will remain, in some particular state. At random intervals the series of messages is interrupted and *S* is asked what the present state of one of the variables is.

The problem considered here is how the fraction of the questions answered correctly depends on the number of states that each variable may assume. Keeping track of variables that have the same set of possible states has been shown to be different from keeping track of variables that have their own exclusive states (Yntema & Mueser, 1960); so it was decided that the effect of number of states should be examined in both situations. It was also decided that the probability that a message would change the variable to which it referred should be held constant while the number of possible states was varied. Preliminary experiments had shown that the task is especially difficult when a large fraction of the messages are changes.

METHOD

Three experiments were done to investigate the effect of number of states in a variety of cases. The experiments were much alike in plan, differing chiefly in the number of variables whose states *S* was to remember, and in whether the variables had the same or different possible states.

Experiment I

In Exp. I *S* attempted to keep track of six variables that had the same set of possible states.

Conditions and design.—Looking ahead at Fig. 1 will be helpful in remembering the conditions considered in the various experiments. As the top panel of Fig. 1 implies, Exp. I included 10 conditions: 2, 3, 5, 8, and 21 possible states, every message being a change; and 2, 3, 5, 8, and 21 states, the probability of change being one-fourth. Eight *S*s each made one run under every condition. The design governing the order in which conditions were presented to *S*s was balanced and was an incomplete latin square. Lists of stimuli for different *S*s were generated independently, and so were lists with different probabilities of change; but the lists presented to a given *S* in the conditions involving a given probability of change resembled each other in a way that will be explained below.

Lists of stimuli.—The variables were identified by the digits 1 through 6. In any one run they all had the same set of possible states, namely, n randomly chosen consonants. The list of stimuli for a run had three parts: the prestarters, the starters, and the body. The prestarters consisted of one message about each variable and were chosen by putting the variables in a random order and then choosing one of the n states for each at random. The starters also consisted of one message about each variable. The variables were again put in a random order and a state was chosen for each, but the probability of choosing the state that had been assigned to the same variable in the prestarters was now $1 - c$, where c was the probability of change, and each of the $n - 1$ other states had probability $c/(n - 1)$. The body of the list consisted of 128 items chosen one after another according to the following scheme: First, it was decided whether the item would be a message or a question. The probability of a question was one-half. Next, the variable to which the item would refer was chosen, all variables being equally likely. Finally, if the item was to be a message a state was selected. The probability that the state would be the one named in the previous message about the chosen variable was $1 - c$, and the probability of each other state was $c/(n - 1)$.

---

[1] Operated with support from the United States Army, Navy, and Air Force.

There is some evidence that randomly generated lists of this sort differ in difficulty (Yntema & Mueser, 1960, p. 21). In order that such differences should not obscure the effect of number of states, the lists presented to a given *S* in the conditions involving a given probability of change were, unknown to *S*, made as nearly alike as possible. To be exact, the variables in any two such lists stood in one-to-one correspondence to each other in the following sense: if the *i*th item on one list was a question, then the *i*th item on the other list was a question about the corresponding variable; if the *i*th item was a message that changed a variable, then the *i*th item on the other list was a message that changed the corresponding variable; and if the item was a message that left the variable unchanged, then the corresponding item was a message about the corresponding variable leaving it unchanged. To reduce the possibility that *S* would learn the common structure of the lists, no variable was ever allowed to correspond to itself.

The lists were generated on an IBM Type 709 computer.

Apparatus and procedure.—The *S* sat before an inclined board to which were fastened six boxes, which represented the six variables whose states he was to remember. The boxes were labeled 1 through 6, and each contained a stack of blank 3 × 5 in. cards.

Before a run *S* wrote down the *n* possible states (unless all 21 consonants were possible) and posted them on a corner of the board where he could refer to them if he wished. He was told that messages would "almost always" change the boxes (if the probabilty of change was one) or would "usually not" change the boxes (if the probability was one-fourth). The *E* then read the prestarters. As each was read *S* wrote the consonant it contained on the top card in the appropriate box and put the card on the bottom of the stack out of sight. He was told he need not try to remember the prestarters.

After those preparations the task began. The *E* read the remaining items in the list one by one. Messages were spoken in the form, "Four B," and questions in the form, "Letter of Two?" When *E* read a message *S* repeated it aloud, wrote the consonant it contained on the top card in the appropriate box, and put the card on the bottom of the stack. As soon as the card was out of sight *E* read the next item. When *E* read a question *S* attempted to reply with the last consonant he had written in the box about which the question inquired, guessing if necessary. He was not told whether his response was correct. As soon as he replied *E* went on to the next item.

The *S* was instructed to work rapidly and not to "take the time to worry about the other boxes" while writing a message or answering a question. There were two runs to a session with a rest of about 25 min. between sessions. On Day 1 there were a practice session and two experimental sessions; the remaining experimental sessions were on Day 2. Before the practice session *E* and an assistant gave a short demonstration of the task. The practice lists were those used in Exp. II and will be described below.

The present method differs from that used earlier (Yntema & Mueser, 1960) in that *S* was prevented from seeing the old state of the variable when writing a message. The present arrangement seemed a little neater in an experiment directly concerned with the probability that the message would be a change; otherwise the difference in method is not considered important.

Subjects.—The *S*s of Exp. I and II were 8 male Air Force lieutenants and 8 Air Force enlisted men. Four of each were chosen at random for Exp. I and the others were assigned to Exp. II. All 16 were operators of an experimental air-defense center. One lieutenant in Exp. I had been an *S* several months earlier in a similar experiment (Yntema & Mueser, 1960). All others were naive to experiments of this type.

Experiment II

In Exp. II there were again six variables (i.e., six boxes), but each now had its own set of *n* possible states. The description of Exp. I applies except where otherwise stated.

There were eight conditions, as may be seen in the middle panel of Fig. 1. The design for the order in which conditions were presented to *S*s was a complete latin square.

The states of the first variable were called Directions. There were at most eight of them—north, northeast, east, etc. They were represented by arrows when *S* wrote them on the cards. The states of the other five variables were: Jewels—diamond, emerald, jade, opal, pearl, ruby, topaz, zircon—represented by their initial letters, written as capitals. Marks—check, plus, hat, stroke, comma, equals, star, question—"hat" represented by a caret, "stroke" by a line slanting down toward the right, "star" by three crossed lines, and the others by the usual symbols. Animals—bird, fish, cow, snake, giraffe, snail, swan, butterfly—represented by crude drawings of the type used earlier (Yntema & Mueser, 1960). Countries—Austria, China, France, Germany, India, Mexico, Norway, Spain—represented by their initial capital letters. Numbers—1 through 8, written as digits. For a run in which every variable was to have n possible states, a subset of size n was drawn at random from each of these six sets. Before the run S wrote the symbols in the six subsets and posted them where he could refer to them if he wished.

The six boxes were labeled "Direction," "Jewel," "Mark," "Animal," "Country," and "Number." They were also labeled with a letter: K, H, L, Q, W, or X. The letter was changed from run to run, but was the same for all boxes in any one run. Its only purpose was to allow the messages and questions to be spoken in a form comparable to that used in practice and in the other experiments—i.e., in the form, "K north," and "Jewel of K?"

All *S*s of Exp. I and II were given the same two lists for practice. The first was a half-length list with four variables: Animal and Country of K and X. The second was full-length and there were nine variables: Jewel, Mark, and Number of H, L, and W. In both lists every variable had four states and the probability of change was three-fourths. The *S*s of the present experiment spent about 5 min. at the start of the practice session writing the 48 symbols and learning to associate them with the names of the states.

Experiment III

Except that there were three variables (i.e., three boxes) instead of six, Exp. III was essentially a repetition of the first two, and the description of Exp. I and II applies here except when otherwise stated. Some of the conditions examined in Exp. I and II were omitted, however. It was felt that if the probability of change were only one-fourth, the task of keeping track of three variables with different sets of states would be so easy that there would not be enough errors to analyze. Then too, keeping track of three variables that have the same two possible states is peculiar: S need only remember which variable if any is the odd one and what its state is. Those conditions were also omitted, since this was not intended as an experiment on variables whose states fall into patterns.

Twelve conditions remained, as may be seen in the bottom panel of Fig. 1. The states were Directions, Marks, and Animals when each variable had its own set. Lists for conditions that differed only in number of states resembled each other as in Exp. I and II, but with only three variables it was no longer possible to require that no variable should ever correspond to itself. The design for the order of conditions was a balanced incomplete latin square.

There were 8 *S*s, all of whom had served in Exp. I or II about a month earlier. There were a lieutenant and an enlisted man from Exp. I, and four lieutenants and two enlisted men from Exp. II. They were chosen solely because they were still available. There was only one practice run, a full-length run with nine variables: Direction, Mark, and Animal of H, K, and L.

RESULTS

Primary results.—The fraction of the questions answered correctly in each run was corrected for chance in the usual manner—that is, by solving for p the equation:

$$r = p + (1 - p)/n \qquad [1]$$

r is the raw fraction right.

n is the number of states each variable may assume.

p is the fraction right after correction for chance.

The correction was used because it made the results easier to describe—not because of any assumption that S guesses when he does not know the answer. It will be shown below that errors are not always random guesses.

Figure 1 shows the mean p for each condition. It may be seen that in none of the seven cases examined in these three experiments does the number of states have any large effect on the corrected fraction right. In fact, as far as these data can show the corrected fraction right is constant. Analyses of variance and nonparametric tests based on the coefficient of concordance (Kendall, 1955, Ch. 6) failed to show that the effect of number of states was significant at the .10 level in any of the seven sets of conditions.

Inspection of Fig. 1 makes it obvious that the probability that a message will change the variable to which it refers has a large effect. As

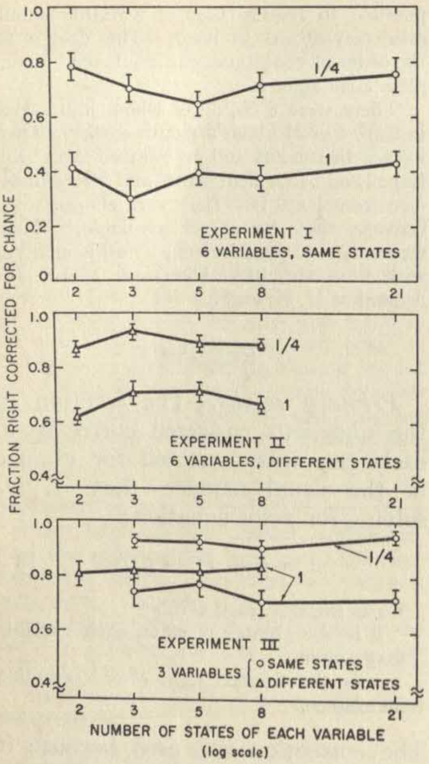


FIG. 1. Effect of *n*, number of states each variable may assume, on fraction of the responses correct after allowance for chance success. (The parameter is the probability that a message will be a change. Points are means for 8 *S*s. Vertical bars extend $\pm\sigma_m$.)

expected, *S*s keep better track of slowly changing situations than of rapidly changing ones.

Secondary results.—The corrected fraction right increased markedly from run to run in Exp. I, the average increase being about .02 per run with either probability of change. It was feared that the effect of number of states might be hidden by this order effect; so the linear component of the order effect was removed from the corrected fractions right observed in Exp. I and the tests for effect of number of states were repeated. The results were the same. The order effect helps to account for the unusually large standard errors shown in the top panel of Fig. 1, but its influence on the shape of the two functions shown there may be dismissed as negligible. Since the order of conditions was balanced, the linear component had no influence. There was no obvious nonlinear component; and since the design was eight-tenths of a latin square, only a quite obvious nonlinear component would have changed the ordinate of any point by as much as .01 relative to the grand mean.

An analysis was made of the frequency with which *S* responded with the previous state of the variable in question instead of its present state. Questions that were answered correctly were ignored. So were questions whose answers had been defined by a message in the starters or by a message that was not a change. The fraction of the remaining questions answered with the previous state was computed for each run with *n* greater than 2 and was corrected for chance by the formula obtained when $n - 1$

TABLE 1

CORRECTED PROBABILITY OF A RESPONSE WITH THE PREVIOUS STATE, GIVEN AN ERROR AND GIVEN THE LAST MESSAGE WAS A CHANGE

Type of Task and Probability of Change	Mean	σ_m
Six Variables:		
Same states, 1/4	.48*	.14
Same states, 1	.04	.04
Different states, 1/4	.52*	.14
Different states, 1	.57**	.08
Three Variables:		
Same states, 1/4[a]	.48*	.14
Same states, 1	.10	.06
Different states, 1	.58**	.08

Note.—Means are for 8 *S*s (except as noted), chance correction applied.

[a] 6 *S*s; the other 2 made no errors when the last message was a change.

* Different from zero at .02 level, two-tailed.

** Different from zero at .001 level, two-tailed.

is substituted for n in Equation 1. The corrected fractions did not appear to depend on n; so they were averaged over the runs made by a given S under conditions differing only in n. The means of the resulting averages are shown in Table 1. It may be seen that in five of the seven cases there was a significant tendency for S to respond with the previous state when he made an error. Why the other two cases should be exceptions is not entirely understood.

DISCUSSION

These experiments agree with previous work on short-term recall of random, independent items (Gregg & Laughery, 1960; Lloyd, Reid, & Feallock, 1960; Miller, 1956; Pollack, 1953; Pollack, Johnson, & Knaff, 1959). Like the earlier results, the present ones can be interpreted as showing that the size of the set of alternatives from which the items are drawn has a relatively small effect on the average number of items that can be retained. In all of the previous experiments the data were collected or analyzed in such a way that chance successes probably did not have much effect on the results.

The improvement from run to run in Exp. I seems to be unusual: no important effects of order were observed in Exp. II or in an earlier experiment on naive *S*s (Yntema & Mueser, 1960). It may have been a mistake to use digits for the names of the variables in Exp. I and consonants for their states. Combinations of digits and consonants are often meaningful: "2B" sounds like "to be," "4F" is a Selective Service classification, and many other pairs have special associations for an Air Force man. With practice the *S*s may have learned to keep track of mediating responses like "to be" instead of keeping track of the consonants. If so, it is understandable that performance improved. The task is easier when the variables have different states (Yntema & Mueser, 1960); and a mediating response, being peculiar to one of the digits, is like a state that can be assumed by only one of the variables.

SUMMARY

The task was to remember the present states of several variables. The S was read a series of messages, each telling him what the state of one of the variables was. He recorded the messages in such a way that he could not see them once he had written them. At random intervals the series of messages was interrupted and he was asked to recall what the last message about one of the variables had been.

The effect of the number of states that each variable might assume was examined in seven cases that differed from each other in one or more of the following ways: (*a*) there were three variables for S to keep track of at once, or there were six; (*b*) all variables had the same set of possible states, or each had its own distinct set of states; and (*c*) the probability that a message would change the state of a variable was one, or it was one-fourth. In none of the seven cases did the number of alternative states have any significant effect on the fraction of the questions answered correctly, provided the fraction correct was adjusted for chance success. The probability of change was found to have a large effect.

REFERENCES

GREGG, L. W., & LAUGHERY, K. R. Memory span for digits as a function of size of symbol set and mode of presentation. *Amer. Psychologist*, 1960, **15**, 478. (Abstract)

KENDALL, M. G. *Rank correlation methods*. (2nd ed.) New York: Hafner, 1955.

LLOYD, K. E., REID, L. S., & FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

POLLACK, I. Assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421–435.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

YNTEMA, D. B., & MUESER, G. E. Remembering the present states of a number of variables. *J. exp. Psychol.*, 1960, **60**, 18–22.

(Received January 23, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 396–403

RECENCY AND FREQUENCY IN PAIRED-ASSOCIATE LEARNING[1]

LLOYD R. PETERSON, DOROTHY SALTZMAN, KENNETH HILLNER, AND VERA LAND

Indiana University

The assumption that repetitions build up associative strength by the accumulation of successive increments has recently been challenged (Estes, 1960; Estes, Hopkins, & Crothers, 1960). A single presentation of a list in experiments by these investigators appears to have produced little or no strength for some pairs of associates, while other pairs acquired high strength. The data have been interpreted as evidence for some form of all-or-none theory of the formation of associations.

An investigation of the time dimension may clarify some aspects of the formation of associations. The effect of a presentation must be assessed by some later measurement, and the interval elapsing between presentation and test may well be related to the number of correctly recalled associations. It seems likely that when perceptual conditions are favorable the probability of a correct recall might be close to 1.0 immediately following presentation. If a gradual decrease in the retention curve also occurs, a reasonable interpretation would be that associations are always formed but differ as to degree of retention.

An attempt to relate short-term retention to learning has been made by measuring recall of individual nonsense items after a single presentation over intervals up to 18 sec. (Peterson & Peterson, 1959). Recall was found to decrease very rapidly, with repetition significantly improving retention. This retention of items over short intervals would seem to relate directly to what Underwood and Schulz (1960) have called the "response recall" stage of list learning, during which time the items comprising the list are learned. Concurrent with this response learning in the list experiment, associations between stimuli and responses are established. The latter kind of cue learning has been investigated in a series of short-term retention experiments to be reported in this paper. Response learning was minimized by specifying in the instructions a set of numbers which were then used as responses repeatedly throughout a given experiment. The initial experiment investigated the retention function for associations over intervals less than the time between trials in the traditional list-learning experiment.

EXPERIMENT I

Method

Subjects and apparatus.—Twenty-four students from the introductory psychology course at Indiana University served as *Ss*. Participation in experiments was a course requirement. A memory drum was set in a 3 × 4 ft. black screen. A sliding shutter permitted exposure of any one of several columns of items. Twenty-eight exposures comprised a complete revolution of the drum.

---

[1] This research has been facilitated by National Science Foundation Grant G-12917, for which the senior author is principal investigator. The first author is responsible for the design, interpretation, and report of the investigation. The second author assisted in Exp. III and IV. The third author assisted in Exp. II. The fourth author assisted in Exp. I.

Procedure.—A set of 108 three- and four-letter single-syllable words occurring at least 10 times per million in Thorndike and Lorge's (1944) count provided the stimuli. The responses were the digits 1–9.

The experimental session was divided into 12 blocks of presentations, each block containing 26 exposures of items, with a 4-sec. rest between blocks, save that between Blocks 6 and 7 a rest interval of 1 min. permitted change of the drum tape. A sample block is shown in Table 1. Nine words were selected at random from the previously described set to be used in a given block of exposures. A digit was randomly paired with each word. Five of these pairs of items were used to test the five experimental conditions, while the other four pairs were used as dummies to fill space. The first exposure in each block was a dummy to eliminate the advantage of initial position. Dummy pairs were presented and repeated where necessary to fill out blocks. The two members of a pair appeared together during a 2-sec. exposure, and recall was tested by presenting the stimulus alone after zero, one, three, or eight other exposures had occurred. Thus retention intervals of 0, 2, 6, and 16 sec. were tested. A repetition condition was also tested by presenting a pair twice in succession and then presenting the stimulus alone 6 sec. after the second presentation of the pair. For all conditions presentation and testing of one condition overlapped the retention cycle of other conditions. Pairs were randomly assigned to conditions within each block. A given word was used in only one block of the experiment, while the same set of digits was used again with new words in each succeeding block. The position of a condition within the block was varied randomly from one block to another. The *S* was instructed to pronounce aloud everything that appeared in the aperture. When only a word appeared, *S* was also to try to speak aloud the number last seen with the word. The *S* was instructed to guess when necessary.

TABLE 1

SAMPLE SEQUENCE IN EXP. I

Exposure	Function	Exposure	Function
HILL 7	dummy	ART 3	dummy
BOAT 9	presentation	YARD	test (6 sec.)
CAR 6	dummy	CAR 6	dummy
YARD 4	presentation	BOAT	test (16 sec.)
YARD 4	repetition	LOT 2	presentation
LINE 1	presentation	HILL 7	dummy
LINE	test (0 sec.)	LOT	test (2 sec.)

Results and Discussion

A single pairing of a word with a number resulted in proportions of recall of 1.00, .76, .40, and .24 after intervals of 0, 2, 6, and 16 sec., respectively. Immediate recall was not quite perfect since one error occurred. Recall decreased rapidly, but even at 16 sec. was well above the .11 chance level. A Friedman two-way analysis of variance resulted in a χ_r^2 of 40.9 which with 3 *df* is significant at the .001 level.

The proportion of recalls after 6 sec. for the two-presentation condition was .55. Seventeen *S*s recalled more after two presentations than after one, while only 3 *S*s recalled more after one presentation. Hence the difference in recall resulting from differing numbers of presentations was significant at the .01 level by two-tailed sign test.

The initial experiment indicates that retention of an association decreases over intervals less than those occurring between trials in list learning. It rarely happened that an association was not retained for any time at all. Any test of the number of associations produced by a single pairing would be influenced by the interval between presentation and test. In Exp. II a successive test procedure was utilized with varying intervals between Presentation, Test 1, and Test 2.

EXPERIMENT II

Method

The techniques of Exp. I were used with minor modifications. A memory drum taking a 32-in. tape was introduced, eliminating the necessity for changing the tape during an experimental session. Table 2 shows a sample block of exposures, the stimuli being trigrams

TABLE 2

Sample Sequence in Exp. II

Exposure	Function	Exposure	Function
DAY 5	dummy	MAY 4	dummy
TAG 1	presentation	TAG	second test
BID 8	presentation	HAY 12	presentation
MAY 4	dummy	CAN 2	dummy
TAG	first test	BID	second test
DAY 5	dummy	ROT 7	presentation
CAN 2	dummy	ROT	first test
BID	first test		

of 99–100% Archer (1960) association value, and the responses being the numbers 1–15. Blocks varied in length from 31 to 35 exposures. A trigram and a number were presented together initially and tested by presenting the trigram alone on two later exposures. Interval A from presentation to Test 1 was either 0, 4, or 8 sec., while Interval B from Test 1 to Test 2 was either 4 or 8 sec. Each of the six combinations of these factors was tested 12 times with each of 36 introductory students. As in Exp. I, the presentation and tests of one condition overlapped that of other conditions. Three tapes assigning different random pairings of numbers and trigrams to the various conditions were made and each was used with 12 *S*s.

Results

The mean proportions of numbers recalled on Test 1 were .99, .37, and .30 for the 0-, 4-, and 8-sec. conditions, respectively. Recall after a single pairing decreased more rapidly than in Exp. I, perhaps as a function of the increased number of alternative responses in Exp. II, where the chance level dropped to .07.

The analysis of recall on Test 2 is presented in Table 3. In all conditions, probabilities were calculated for individual *S*s and means computed from these scores. The proportions recalled were inversely related to the length of both Intervals A and B. An analysis of variance showed that both Intervals A and B produced *F*s significant at the .001 level, using their respective interactions with *S*s as the error terms ($F = 14.9$, $df = 2/70$; $F = 14.0$, $df = 1/35$). The AB interaction was also significant at the .001 level, using the triple interaction mean square as the error term ($F = 12.3$, $df = 2/70$). Recall on Test 2 was thus a function of the total elapsed time since presentation, and the usual finding of most rapid forgetting early in the interval immediately after presentation was also found here.

The comparisons of greatest interest are among the conditional probabilities of recall on Test 2 given a particular outcome on Test 1. The probability of a correct response on Test 2 given a correct response on Test 1 increased as the length of Interval A increased and decreased as Interval B increased. An analysis of variance found both these factors to be significant at the .001 level

TABLE 3

Recall on Test 2, Exp. II

Interval B	Proportion Recalled			Probability of Correct Given Correct on Test 1			Probability of Correct Given Incorrect on Test 1		Probability of an Error Given that Error on Test 1	
	Interval A			Interval A			Interval A		Interval A	
	0 Sec.	4 Sec.	8 Sec.	0 Sec.	4 Sec.	8 Sec.	4 Sec.	8 Sec.	4 Sec.	8 Sec.
4 Sec.	.48	.34	.24	.49	.66	.71	.14	.06	.42	.49
8 Sec.	.30	.27	.25	.30	.46	.52	.29	.10	.28	.28

($F = 15.34$, $df = 2/70$; $F = 30.6$, $df = 1/35$), while their interaction F was less than 1. The interpretation of the effect of Interval B on cases where S is correct on Test 1 seems straightforward. In this case the beginning of the retention interval may be taken to be the removal of the test trigram from the drum aperture. Then the probability of recall decreases as the interval lengthens, in accord with the previously established functional relationship between recall probability and time. The increase in the dependent probabilities as Interval A lengthens suggests a selection of items. Those associations strong enough to survive to Test 1 over the longer intervals are those most likely to endure to Test 2. An alternate explanation would be that assuming Test 1 is a second presentation for pairs correctly recalled, then distributed presentations may be retained better than massed presentations. Either interpretation would seem to imply a continuum of associative strength.

The conditional probabilities of a correct response on Test 2 given an incorrect response on Test 1 decreased as Interval A increased, while these probabilities increased as Interval B increased. The 0-sec. conditions of Interval A are omitted because of the scarcity of incorrect responses in these conditions. An analysis of variance of the remaining conditions showed that Interval A was significant at the .001 level ($F = 8.81$, $df = 2/70$), while Interval B was significant at the .05 level ($F = 4.58$, $df = 1/35$). The interaction F was less than 1. A decrease in conditional probability as Interval A increases is understandable if Test 1 strengthens an interfering association. Test 1 would be expected to have greater relative influence the farther it is removed from the original presentation. The direct relationship between recall and the length of Interval B is also consistent with this interpretation. As time since pairing of cue and interfering response passes, the strength of this association decreases with the result that the relative strength of the original association increases. Thus a kind of spontaneous recovery of strength occurs that is in line with findings in retention experiments over longer intervals (Briggs, 1954). Strong evidence of interference from incorrect responses on Test 1 appeared in a comparison of the two conditions with equal time from presentation to Test 2. In the condition where Test 1 was closer to Test 2 than to presentation, the probability of a correct response on Test 2, given incorrect on Test 1, was markedly lower than in the other condition.

The strengthening effect of a test without knowledge of results may be inferred from the data on responses which were wrong on Test 1 and repeated on Test 2. Thirty-three *S*s made wrong responses in all four conditions. The conditional probabilities of the same error of commission being made on Test 2 as on Test 1 were computed for these *S*s and summarized in Table 3. A Friedman nonparametric two-way analysis of variance showed these to be significantly different at the .001 level ($\chi_r^2 = 16.6$, $df = 3$). The decrease in probability of recurrence as Interval B increased is suggestive of a retention effect following a contiguous pairing of stimulus and response. In this case S has determined the pairing instead of E, but the result seems similar to an E-arranged pairing. Note that the probability of an error recurring after 4 sec. was somewhat higher than the probability of recall of an E-paired response after a corresponding interval, although the difference was not statistically significant. It is to be expected that all of the response alternatives are not equally likely to occur as errors, and therefore a selective factor may have influenced the occurrence of errors.

The data support the hypothesis that a test changes associative strength even though knowledge of results is withheld. When a response is correctly recalled, the test can be interpreted as a second pairing. Not only is the association reinstated at the zero point of the retention function, but the slope of the function is changed. When an error occurs at Test 1, a pairing of the stimulus with a new response sets the stage for retroactive interference. Thus, in successive tests, the probability of a right following a right would be expected to be high, while the probability of a right following a wrong should be low. This interpretation was tested further in Exp. III by having *E* either force a repetition of the original pairing or introduce a new pairing at a point in time equal to that of the first test in a double test condition.

EXPERIMENT III

Method

Four conditions were used with each of 28 *S*s from the introductory psychology course. Condition DT (double test) reproduced the basic paradigm of Exp. II with a 4-sec. Interval A and a 4-sec. Interval B. Condition R (repetition) substituted a second presentation of the stimulus and response for Test 1 of the above condition. Condition I (interference) paired the original stimulus with a new response in place of Test 1 of Cond. DT. Condition S (single) tested retention once 10 sec. after a single presentation. Thus all four conditions were tested for recall 10 sec. after the original pairing. As in Exp. I and II, the conditions overlapped one another in each block.

The *S* was instructed to speak the number that had been paired with the word the first time that he saw it. Stimuli were common three- and four-letter words as in Exp. I. Responses were the numbers 1–12. Each block contained 29 exposures and 8 sec. elapsed between blocks. Each condition was tested twice within a block for eight blocks. Four tapes were prepared such that a particular pairing of word and number appeared equally often in each of the four conditions over the group of *S*s.

Results and Discussion

The mean proportions recalled after 10 sec. in Cond. S and DT were .27 and .28, respectively, indicating that insertion of a test midway in the 10-sec. recall interval did not alter the overall proportion recalled. Any strengthening of associations correctly recalled is apparently offset by strengthening of erroneous responses. The mean proportion recalled in Cond. R was .56, showing that an *E*-controlled repetition did markedly increase recall. A two-tailed sign test was used to test this difference as well as the other comparisons in Exp. III. Condition R was superior to Cond. S for 25 of the 28 *S*s, indicating a difference significant at the .01 level. The proportion recalled in Cond. I was .14, which was a marked decrease relative to Cond. S. This decrease after insertion of a different pairing was found with 21 *S*s, while 3 *S*s showed an increase, and 4 showed no change ($P < .01$).

The conditional probability in Cond. DT that a response was recalled on Test 2, given that it was correct on Test 1, was .68. The superiority of this to the mean proportion recalled in Cond. R was small, but it was found with 21 *S*s ($P < .05$). The conditional probability that a response was recalled correctly on Test 2 given that it was incorrect on Test 1 was .10. This was only slightly lower than Cond. I, but the difference was found with 19 *S*s and only 6 *S*s showed the reverse difference ($P < .05$).

Thus the assumption that Test 1 strengthens or weakens the initially presented association depending on whether it is correctly or incorrectly recalled could account for the major part of the spreading of the conditional probabilities in Cond. DT but not all of it.

The presence of a selective factor is suggested by another comparison. The conditional probability in Cond. DT of an error on Test 2, given the same error on Test 1, was .46. This was appreciably higher than the proportion of second-presentation responses recalled in Cond. I, .27. Twenty *S*s showed the difference in this direction ($P < .05$). This suggests either differential initial probabilities of response to given items or differential difficulty of association, and it is only reasonable that a selective factor is also influencing correct responses on Test 1. Considering the joint evidence of Exp. II and III, the high probability of a correct recall on Test 2 given a correct recall on Test 1 appears to be due to the combined effect of selection and strengthening on Test 1. Individual differences among items are presumably related to past histories of *S*s. The effect of the test is to make the strong stronger and the weak weaker.

EXPERIMENT IV

The equivalence of a pairing read from the drum and one not so prompted was further investigated by testing short-term retention of an association after two pairings under several conditions of elicitation of the second pairing.

Method

Four conditions were used with each of 30 *S*s from the introductory course. Three conditions of repetition differed as to manner of production of the second pairing. The first pairing was produced by exposure of both stimulus and response in the aperture. In Cond. SR-P (SR-Prompt) both stimulus and response were shown again immediately after the first presentation. In Cond. S-P (S-Prompt) only the stimulus was shown on the exposure following the first presentation. In Cond. NP (No Prompt) neither stimulus nor response was shown *S*, but a set of ditto marks appeared following the first presentation. The *S* was instructed to read everything that appeared in the drum aperture, to supply the missing response whenever the stimulus appeared alone, and to repeat both stimulus and response just seen when the ditto marks appeared. An example of each type of presentation was shown *S*s on cards during the instructional preparation.

Recall was tested in each of these conditions by presenting the stimulus alone 6 sec. after the second presentation. In Cond. S (Single) recall was tested 6 sec. after one presentation.

The stimuli were trigrams of 99–100% Archer value. The responses were the numbers 1–10. Blocks were 26 exposures long, and 8 sec. elapsed between blocks. The three conditions of repetition were tested twice per block over 18 blocks, while Cond. S was tested once per block over 18 blocks. Conditions overlapped one another in each block. The order in which items were tested was randomized within each block. Three randomized assignments of pairs to conditions were made and each was used with 10 *S*s.

Results and Discussion

The mean proportions recalled were .42 ($\sigma_m = .02$) for Cond. SR-P, .44 ($\sigma_m = .02$) for Cond. S-P, .40 ($\sigma_m = .02$) for Cond. NP, and .35 ($\sigma_m = .03$) for Cond. S. The distributions of proportions for individual *S*s appeared to be approximately normal, and therefore a parametric test of significance was used. Conditions SR-P and S-P were both significantly different from Cond. S at the .01 level ($t = 3.1, 4.0$; $df = 29$), but did not differ significantly from each other. Condition NP was not significantly different from Cond. S, but was different from Cond. S-P at the .05 level ($t = 2.25$). The inferiority of Cond. NP was not due to failure of *S*s to respond appropriately to the ditto marks, since no such failures were observed. The *S*'s behavior on second presentation was identical in all three conditions. Recall was well above the chance level of .10 for all conditions.

It may be concluded that *S*'s production of the response in the presence of the stimulus alone strengthens the association at least as much as reading the response from the drum. The difference in the two conditions in which the response was not prompted could perhaps be predicted from considerations of stimulus generalization. In Cond. S-P the stimulating situation is more similar to that obtaining in test for recall than is the case in Cond. NP.

The possibility of a progressive deterioration in performance over the experimental session (proactive interference) was investigated by summing correct responses in the three repetition conditions within successive combinations of three blocks. The proportions correct in successive stages were .42, .36, .44, .42, .43, and .46. A simple analysis of variance resulted in an *F* of 3.03 which was significant at the .05 level ($df = 5/145$). A *t* test between the first and last stages did not show a significant difference ($t = 1.34$, $df = 29$), and in any event the difference was in the wrong direction to be evidence for PI. There might be a transient interference effect in the second stage, but it is not marked. Progressive effects were further investigated by going back to the data of Exp. III and summing number correct after 10 sec. for all four conditions over successive pairs of blocks. The proportions obtained for the successive stages were .31, .28, .34, and .33. The *F* obtained in a simple analysis of variance was not significant, but again there is a slight dip in the second stage analyzed. Again, too, there is the possibility of a small practice effect from beginning to end of the session. In all the experiments described, any progressive effect was controlled by the equal appearance of all experimental conditions in each stage of the experiment.

CONCLUSIONS

It has been argued that short-term retention is qualitatively different from learning and long-term retention (Clark, Lansford, & Dallenbach, 1960). However, this study and an earlier one (Peterson & Peterson, 1959) have found some of the same phenomena in short-term retention that had previously been found in learning and retention studies over longer intervals, for example, increased recall through repetition, retroactive interference, and spontaneous recovery. It would seem parsimonious to assume that short-term retention is basically similar to retention over longer intervals until experimental evidence dictates otherwise.

The results of the four experiments taken together suggest that any interpretation of the successive test experiment must account for both the selection of items in relation to the passage of time and the effect of the first test in producing another pairing of the cue with some response. The change in the conditional probabilities on the second test as a function of the time between presentation and first test would seem to be a particularly recalcitrant characteristic for an all-or-none theory to accommodate. The results would appear to be most readily interpreted within the framework of an incremental theory of learning.

SUMMARY

The formation of associations between pairs of verbal items was investigated in four short-term retention experiments. Experiment I showed retention to decrease regularly over a 16-sec. interval, although recall was nearly perfect immediately after presentation. In Exp. II the effect of a test without knowledge of results (Test 1) was found to vary with its position between original presentation (P)

and a second test (Test 2). Correct recall on Test 2, given correct recall on Test 1, increased as the time between P and Test 1 increased; correct recall on Test 2, given incorrect recall on Test 1, decreased as the time between P and Test 1 increased; error on Test 2, given that error on Test 1, decreased as the interval between tests increased. In Exp. III it was found that the strengthening or interference which resulted from substituting forced pairing of the same or different responses for Test 1 could not completely account for the diverging conditional probabilities in a successive test experiment. In Exp. IV the increase in recall from an unprompted response without knowledge of results was found to be as great as that for a prompted repetition.

The findings are discussed in relation to the all-or-none vs. incremental assumption regarding associative learning, and are interpreted as favoring the latter assumption.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74** (10, Whole No. 497).

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

CLARK, L. L., LANSFORD, T. G., & DALLENBACH, K. M. Repetition and associative learning. *Amer. J. Psychol.*, 1960, **73**, 22–40.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received January 26, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 404–408

A TEST OF A GENERAL UTILITY THEORY MODEL FOR PROBABILITY LEARNING[1]

KENNETH B. LITTLE
University of Denver

YVONNE BRACKBILL AND STEPHEN H. KASSEL
Johns Hopkins University

The model presented in this report is an extension to the multichoice case of Siegel's (1959) Model I for probability learning in the two choice situation. The extended model asserts that in a noncontingent probability learning situation, the asymptotic response level of *S* will be a function of the relative frequency of occurrence, π_i, of the stimulus event within the set of stimulus events, of the marginal utility of making a correct response, *a*, and of the marginal utility for *S* of varying his responses, *b*.

In the general case of *k* alternatives, the expected utility of a correct prediction (U_r) is:

$$U_r = a \sum_{i=1}^{k} p_i \pi_i \qquad [1]$$

where *k* is the number of choices, and p_i is the proportion of correct responses for the event appearing with relative frequency π_i.

The expected utility of varying responses (U_v) is defined as *b* times the sum of the variances of p_i:

$$U_v = b \sum_{i=1}^{k} [p_i(1 - p_i)] \qquad [2]$$

This function has a maximal value when all $p_i = 1/k$, i.e., when *S* distributes his responses equally over all alternatives.

In order to find that value of *p* that maximizes total expected utility (*U*), it is necessary to solve the $k - 1$ simultaneous equations resulting from the $k - 1$ partial derivatives of the equation for *U*:

$$U = U_v + U_r = a \sum_{i=1}^{k} p_i \pi_i + b \sum_{i=1}^{k} [p_i(1 - p_i)] \qquad [3]$$

The solution is:

$$p_i = \frac{1}{k} + \frac{a}{b}\left[\frac{k\pi_i - 1}{2k}\right] \qquad [4]$$

When π_i and *k* are fixed the model indicates that *S*'s strategy will be determined by the ratio of the marginal utilities, *a*/*b*. The value of this ratio, α, can be estimated from any set of probability learning data by the equation:[2]

$$\hat{\alpha} = 2\left[\frac{p_i - \frac{1}{k}}{\pi_i - \frac{1}{k}}\right] \qquad [5]$$

There is no a priori reason to assume that α will vary as a function of *k* or π. Consequently, once the α value for *S* is determined from one set of π and *k* values it should be possible to predict *S*'s stable state response proportions for any other set of π and *k* values. As a test of

---

[1] This study was done while the second author held a United States Public Health Service Research Fellowship.

[2] This estimate differs from Siegel's (1959, p. 308) for the two-choice case by a factor of 2. The difference comes from using the sum of the variances rather than the average variance in the equation for U_v.

the model, the following experiment was conducted.[3]

TABLE 1

Mean Response Proportions for the Final 100 Trials

Study	Cond.	N	Measure	Choices			
				1	2	3	4
Basic	A	10	π	.500	.250	.250	
			p_o	.429	.299	.272	
	B	10	π	.500	.300	.200	
			p_o	.438	.302	.260	
	C	10	π	.600	.200	.200	
			p_o	.577	.253	.170	
	D	10	π	.600	.250	.150	
			p_o	.588	.211	.201	
	E	10	π	.500	.167	.167	.167
			p_o	.431	.178	.203	.188
	F	10	π	.500	.250	.150	.100
			p_o	.431	.235	.170	.164
	G	10	π	.600	.133	.133	.133
			p_o	.590	.134	.137	.139
	H	10	π	.600	.200	.150	.050
			p_o	.570	.156	.163	.111
Sup. A	E to C	8	π	.600	.200	.200	
			p_o	.615	.225	.160	
	B to G	8	π	.600	.133	.133	.133
			p_o	.561	.145	.148	.146
Sup. B	B	9	π	.500	.300	.200	
			p_o	.443	.271	.286	
	D	9	π	.600	.250	.150	
			p_o	.599	.231	.170	

Method

Subjects.—The *S*s were 80 sixth grade students selected at random from classes within a single elementary school. The mean age was 12.0 yr. The *S*s were assigned randomly to eight experimental groups, with the single restriction that equal numbers of boys and girls be included in each group. The groups represented different combinations of π and k values, the two variables of principal concern in this experiment. Also represented was a third variable of minor importance, even vs. uneven splits in the frequencies of occurrence of the less frequently occurring events. The eight conditions—together with the results—are presented in the upper part of Table 1.

Experimental material.—The stimuli were 11,200 3 × 5 in. colored cards. On one side of each card was printed a 1-in. high letter. The four letters used, E, T, A, and O, were chosen for their high frequency of occurrence in written English. Twenty-eight packs of 400 cards each were prepared for the eight conditions. For each pack of 400 cards, randomized series were prepared (and recorded) in blocks of 20 trials each. The particular objective probability split needed for any condition was maintained within each block (or approximately so, in the case of Treatments E and G). There were no other restrictions on randomization. Since the blocks were interchangeable, it was possible to make up a separately randomized order for each *S* in each group.

Procedure.—The *S*s were tested individually in an experimental room in the school. The task was explained to *S*, and he understood that he was to guess the letter on the back of each card before it was turned over and to make as many correct predictions as possible. Trial 200 was followed by a rest period of approximately 2 min. Following Trial 400, *E* asked *S* to estimate the frequency of occurrence of each letter.

Results and Discussion

A preliminary analysis of variance for each of the eight groups indicated no significant differences among the means of the five blocks of trials beginning with Trial 301. By this criterion stable response levels had been reached. Table 1 shows, for each condition of the experiment, mean response proportions based on these final 100 trials.

In order to test the accuracy of prediction of the model, individual α scores were computed for each of the 80 *S*s on the basis of their asymptotic response proportions to the most frequently appearing event, π_A. An initial analysis indicated that, as expected, there were no significant differences in the mean α values attributable to even vs. uneven splits in the π values of the less frequently occurring events. Therefore, the data were pooled for Groups A and B, C and D, E and F, and G and H. A two-way analysis of variance was then computed to determine the effects on α of major π value (.500 or .600) and number of choices ($k = 3$ or 4). The analysis indicated that α's based on the two levels of k were not significantly different ($F = .27$), but that there was a significant difference between α's as a function of π value ($F = 6.13, df = 1/76$, $P < .02$). Apparently, the model failed to predict accurately from one π value to the other.

---

[3] The authors are grateful for the cooperation of the Baltimore City Schools and for the help of J. T. Ahrens and Anthony Bravos in collecting the data for Supplementary Study B.

This result was surprising in view of the relative accuracy of the model when applied to various data reported in the literature (Greenhouse, Little, Brackbill, & Kassel, 1960). Absolute differences between asymptotic response proportions obtained by Gardner (1957, 1958) and Cotton and Rechtschaffen (1958) and those predicted by the present model averaged only .029 even when k varied from 3 to 9 and π_A from .60 to .70. Since the major difference (with regard to the parameters of the model) between those studies and the present one was our use of a π_A value as low as .500, it seemed more than possible that the model's failure to predict was somehow associated with *S*s' performance under the .500 π_A conditions. This suspicion was reinforced by the marked and consistent difference in closeness of probability matching between the four .500 π_A groups and the four .600 π_A groups. If the suspicion were true, it would mean that our original assumption was incorrect and that α is dependent on π.

However, there is an alternative explanation. In view of their age level, the possibility existed that *S*s were unable to discriminate adequately among the frequencies of event occurrence in the .500 π_A series. Using the ratio of π_A to π_B (the next most frequently appearing event) as an index of discriminability, then on the average, the .500 π_A series presented more difficult tasks of discrimination than the .600 π_A series. However, if one examines the data in Table 1, it is clear that there is no systematic relationship between the magnitude of the p_A values and the π_A/π_B ratios, within either the .500 π_A or the .600 π_A sets. The *S*s performed at about the same level when the ratios were small as when they were large. In addition, an examination of *S*s' estimates of the relative frequency of occurrence of the stimuli revealed that their average error of estimate was somewhat larger for the .600 π_A series (.17) than for the .500 π_A series (.09).

Since both lines of evidence suggest that poor discrimination was not the factor responsible for the low response levels to π_A characterizing the .500 π_A series, we are left with the first proposition: that α is not independent of π. With low values of π_A, apparently the marginal utility of correct predictions is reduced, or the marginal utility of varying responses increases. A decrease in U_r is unlikely, in view of the results of a recent experiment by Brackbill, Kappy, and Starr (1962), who found a higher utility of correct predictions for an event of low frequency than one of high frequency.

It appears, then, that the model's failure to predict across π values is attributable to the fact that α *is* a function of π under certain conditions. Specifically, the use of low values of π_A causes a disproportionate increase in the utility of varying responses; this, in turn, lowers the value of α, with the net result that predicted response levels are in error when one moves to higher π_A series.

Two supplementary studies were conducted to test the stability of the present results and the degree of confidence that could be placed in our interpretation of them.

Supplementary Study A.—The purpose of this study was to determine the extent of individual consistency in the utility of varying responses from one experimental condition to another, and also to ensure against the possibility that sampling error had in some unknown way contributed to the original results for the .500 π_A groups.

Eight each of the original *S*s tested under Cond. B and E were rerun under Cond. G and C, respectively, using the same experimental material and procedure as previously. Approximately 6 mo. intervened between testings.

The second part of Table 1 shows the mean response proportions obtained during the final 100 trials for the two groups. Individual α scores were computed from the obtained response proportions to the π_A event. These α scores were then compared, in a two-way analysis of variance, to the α scores of the 20 *S*s who were originally tested under Cond. C and G. (Number of choices, k, provided the second classification.) All values of F—groups, k, and interaction—were less than 1.00, clearly nonsignificant. These results reinforce our confidence in the adequacy of the original sampling procedure, and indirectly support the hypothesis that α is influenced by low values of π.

Despite a test-retest interval of 6 mo., the rank order correlation between the two sets of α scores for the 16 *S*s was .72 ($P < .01$). The ratio of the utility of correct responses to the utility of varying responses is apparently a fairly stable individual difference characteristic. In line with the argument presented earlier, this is more probably a reflection of individual differences in U_v than in U_r.

Supplementary Study B.—Two new groups of *S*s were tested on two occasions under Cond. B and D ($\pi = 50\text{-}30\text{-}20$ and $\pi = 60\text{-}25\text{-}15$). The purpose was twofold: (*a*) as a partial replication of the original experiment on the effects of major π value, and (*b*) to provide additional data on the stability of individual differences in response variability noted in Study A. Eighteen naive sixth grade *S*s were randomly assigned to two groups, which differed from each other only in terms of the sequence in which the two conditions were administered. The experimental material and procedure were the same as those used previously. A 1-wk. interval separated the two sessions for each *S*.

The lower part of Table 1 shows for both groups the mean response proportions obtained from the last 100 trials of the first testing session. Individual α scores were computed as before. A two-way analysis of variance was then used to compare α scores for the present group to the scores of the 20 *S*s run under Cond. B and D in the original experiment. Of the two variance classifications, experiments and π_A values, only the latter is significant ($F = 5.08, df = 1/32$, $P < .05$). As in Study A, there was a significant correlation between test and retest α scores ($\rho = .66, P < .01$).

DISCUSSION

Considering the present experimental results together with those reported earlier (Greenhouse et al., 1960), the extension of Siegel's utility theory model probability learning must be considered only partly adequate. From α scores determined under one set of conditions, one can predict with reasonable accuracy the final response levels for other conditions that differ in the number of alternatives or choices. However, when the frequency of occurrence of the events is changed, prediction is more uncertain. The model is apparently satisfactory when $\pi_A = .600$ or .700 but fails when relative frequency of the major event drops to .500. This suggests that the model as originally proposed may hold only within a limited range of π_A values, and that the addition of a correction factor may be necessary for accurate prediction within a wider range.

It is interesting that *S*s who received the π_A .500 series did not match probabilities in their asymptotic performance as closely as did *S*s who received the π_A .600 series. Since multichoice noncontingent data have not been reported with π_A values of .500 for adults, it is unclear whether the phenomenon is a general one or reflects a developmental trend.

SUMMARY

A utility theory model for probability learning in the multichoice, noncontingent case was derived as an extension of Siegel's (1959) Model I. To test the model, 80 sixth-grade children were tested individually in a probability learning experiment. The *S*s were required to predict the occurrence of alphabetical letters on 400 consecutive trials. There were three experimental conditions: the relative frequency (π_A) of the most frequently appearing letter was either .500 or .600; the number of choices (k) was either 3 or 4; and the less frequently appearing letters were either evenly or unevenly divided in frequency of occurrence. Ten *S*s were tested under each of the eight possible combinations of conditions thus formed.

The results indicated that (*a*) the model predicted asymptotic response proportions satisfactorily across number of choices but not across π values; (*b*) final response values for the most frequently appearing stimulus in the .500 π_A conditions were substantially below probability matching; and (*c*) the even or uneven division of frequency of occurrence of less frequently appearing stimuli did not significantly affect the response to the π_A letters. Subsidiary studies also demonstrated that α, defined as the ratio of the marginal utility of correct responses to the marginal utility of varying responses, is relatively stable over time for individual *S*s, regardless of changes in π or k values.

REFERENCES

BRACKBILL, Y., KAPPY, M. S., & STARR, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63**, 32–35.

COTTON, J. W., & RECHTSCHAFFEN, A. Replication report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

GARDNER, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

GARDNER, R. A. Perception of relative frequency as a function of the number of stimulus and response categories. *USA Med. Res. Lab. Rep.*, 1958, No. 328.

GREENHOUSE, S., LITTLE, K. B., BRACKBILL, Y., & KASSEL, S. H. A model for multiresponse choice probability learning. *Amer. Psychologist*, 1960, **15**, 456.

SIEGEL, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

(Received February 6, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 409–412

PRESENTATION TIME, TOTAL TIME, AND MEDIATION IN PAIRED-ASSOCIATE LEARNING[1]

B. R. BUGELSKI

University of Buffalo

The present experiment deals with a neglected variable in the systematic study of memorization; namely, item presentation time per trial in relation to total learning time. This variable is obviously important in the interpretation of claims such as that of Rock (1957) for one-trial learning, claims that were based on an interitem time of 8 sec. For Rock, a trial consisted of 96 sec. during which 12 nonsense pairs were exposed. During this period *S* was free to repeat any item for the entire period, but even if he merely used the 8-sec. per item interval, he could still repeat any item quite frequently. The claim that repetition has no merit is obviously improperly drawn from such a study. Time per trial is also involved directly in studies of massed vs. spaced learning where massing is defined as a rapid interitem rate (Hovland, 1938). Using a rate of 2 sec. per item Hovland found learning inferior (14.89 trials) to a rate of 4 sec. per item (6.78 trials). A simple multiplication of trials by time, however, shows the two procedures to involve about the same total time (29.98 and 27.12 sec.). Even when a long intertrial interval is employed, total times appear to be about the same (22.36 and 23.40 sec.) for 2-sec. and 4-sec. presentation rates although the number of trials differed significantly (11.18 and 5.85).

It is the present hypothesis that in at least some areas of memorization, and under some conditions of presentation, the degree of learning will be a function of total time, regardless of the duration of the individual trials or interitem times. The hypothesis does not appear to be without support. Murdock (1960), employing free recall of words in lists of varying length, found total time to be a determining factor in how much was learned regardless of presentation rate. Attempts to compute total learning times from published studies have also proved suggestive for a variety of learning situations but the many variations in kinds and length of material and associated learning conditions prevent adequate comparisons. It is hoped that the present study will focus more attention on the various temporal factors in learning such as exposure time per item, interitem interval, intertrial interval, and total learning time. The essence of the problem is: what is a learning trial?

METHOD

Apparatus.—The Hunter Card Master was used to expose pairs of nonsense syllables at varying presentation times. The stimulus syllable was always exposed for 2 sec. and remained visible while the response syllable was exposed for either 2, 4, 6, 8, or 15 additional sec. Two seconds always elapsed before a new stimulus syllable appeared. Eight pairs of syllables which had been used by Rock, 1957 for his control group (see Table 2), were presented in three different random orders. The total time available to *S* for reacting to or learning a given pair, including the time before a new stimulus was presented,

---

[1] This study was supported by United States Public Health Research and Fellowship Branch Grant M-3489. The writer wishes to acknowledge the aid of N. Carriero and J. Rickwood in gathering the data.

varied, therefore, from 6 sec. for the 2-sec. response to 19 sec. for the 15-sec. response, with the other times lasting 8, 10, and 12 sec. per pair. The time allowed for anticipation of the response term was, of course, always 2 sec.

It will be noted that the total time for presentation of eight pairs of syllables, i.e., one trial, would last either 48, 64, 80, 96, or 152 sec. depending upon the exposure time of the response syllable. There was no rest period between trials.

Procedure.—Each *S* learned the same eight pairs of syllables to a criterion of two successful anticipations of the complete list. The apparatus used did not allow for the elimination of pairs as they were learned, but this was done in the computations so that the total learning times reported are based on the time taken to learn each pair to the criterion. Upon completion of learning each *S* was asked specifically how he happened to learn each pair, what made him think of the correct response, why some pairs were easy and some difficult.

Subjects.—One hundred sophomore students in experimental psychology courses, naive with respect to nonsense syllable learning, were assigned to one of the five experimental conditions, thereby making up five groups of 20. Assignment to groups depended only on order of arrival at the laboratory; the first *S* was assigned to Group 1, the next to Group 2, etc.

RESULTS

Both trials and total time to learn are shown in Table 1 for each of the five different presentation times. The total times are obtained by multiplying the total number of trials to learn each pair in the series by the presentation time. As expected, the longer the presentation time per pair, the fewer the number of trials to criterion. Analysis of variance for the trials yields $F = 18.53$ ($P = .01$, $df = 99$). Total time, however, is not significantly different among the five conditions ($F = 0.845$). The correspondence between total times for the most rapid presentation time and the longest presentation time is striking. Although the former required more than three times as many trials as the latter, the mean time difference is hardly more than 1 sec. Extrapolating roughly from these findings, it might be possible for similar *S*s to learn all eight pairs in one trial with a presentation time of about 25 sec. per pair. Within the limits, then, of the intervals and conditions used, the conclusion appears reasonable that presentation time multiplied by trials amounts to a constant value, $T_p \times T = k$.

TABLE 1

MEAN TRIALS AND TOTAL TIMES TO LEARN LISTS

Presentation Time	*N*	Trials		Total Exposure Time	
		Mean	*SD*	Mean	*SD*
6 sec.	20	10.2	4.1	61.2	24.3
8 sec.	20	8.8	3.8	70.1	30.4
10 sec.	20	5.8	1.9	57.9	19.0
12 sec.	20	4.7	2.5	56.1	29.6
19 sec.	20	3.3	1.2	62.2	22.2

During the postexperimental interrogation every *S* reported spontaneously that one kind of association, mediator or bridge occurred to him and was used to fix the responses. They did not, of course, use such terms and their explanations were ordered into a tentative and arbitrary classification of five seemingly discriminable groups such as the following:

I. The *S* formed one word of the two syllables, as, for example, from DUP-TEZ, he would create DEPUTIZE; from GAC-QET, he would form RACQUET or JACKET.

II. The syllables would be formed into two separate words or a phrase that could be initiated by the stimulus syllable: CEZ-MUN becomes SAYS MAN or C'EST MAN or SEND MONEY.[2]

III. The *S* would manufacture or imagine appropriate phrases or words that sound like the syllables, e.g., GEY-NUR becomes

---

[2] Categories I and II appear to be "sound bound." The associations appear to derive from the sounds of the syllables. The sounds are modified into words.

TABLE 2

Mean Trials to Learn Each Pair of Associates and the Corresponding Number of Mediators Volunteered by *S*s, with Breakdown by Type of Mediator

Syllables	Trials		Total (out of a possible 100)	Mediators					
				Types					
	Mean	*SD*		I	II	III	IV	V	None
GEY-NUR	5.02	5.2	81	54	12	8	2	5	19
KAR-WEH	5.51	5.8	84	7	68	4	3	2	16
BIH-XIR	5.53	4.6	73	18	2	16	29	8	27
CEZ-MUN	6.79	6.2	70	6	31	13	4	16	30
FAX-SOQ	6.94	4.7	57	1	11	26	15	4	43
TOF-LAH	7.13	5.6	54	5	19	7	9	14	46
DUP-TEZ	7.16	5.8	52	2	1	32	5	12	48
GAC-QET	8.24	5.6	66	25	0	15	21	5	34

A GENERAL NURSE or KAR-WEH becomes KAREN-WAYWARD. This category involves largely personal experience.

IV. Some *S*s would attempt abstract analyses of the syllables to get at some lead to the response, e.g., BIH-XIR becomes $BI = 2$, $X = 10$, ONE NUMBER GOES WITH ANOTHER NUMBER, or THEY BOTH HAD AN "I."

V. The *S*s would report some vague association with one part of the combination, e.g., GAC-QET becomes translated into TOURNIQUET, BIH-XIR becomes SOMETHING LIKE BICEPS.

Not all pairs aroused reportable mediators for all *S*s (see Table 2). Some mediators were forgotten after being used. Useless mediators were also commonly mentioned but those were impossible to classify and frequently forgotten. Many *S*s reported inability to learn until a mediator of some type did occur. The longer the interitem interval the more likely the occurrence of a mediator. The five intervals in order from shortest to longest aroused respectively 94, 96, 109, 111, and 127 mediating associations.

Discussion

The findings support the original hypothesis that total learning time is a significant variable to consider in at least some kinds of learning. While it may be convenient to break up a learning session into some kinds of units of time, and label these as "trials," it may be a questionable practice in trying to get at an appreciation of what *S* is actually learning or doing. If a presentation time is anything upwards of 1 sec., he may very well repeat the material to himself within the trial. To count the period as one trial may be quite inappropriate. To conclude, where *S* learns after one such interval, that he learned in one trial (implying no repetition) may be quite incorrect.

On the other hand, the reports from *S*s support the interpretation that on many of the trials nothing effective is learned. Wrong responses may occur because inadequate mediators occurred to *S* or the mediators misled *S* into an inappropriate response. After varying amounts of search, which takes more or less time, an appropriate mediator might be generated or stimulated into functioning. Once this occurs, *S* might be able to respond correctly and meet the learning criterion. In a sense, then, learning does occur in one trial in such situations. It must be remembered that about a third of the time *S* is unable to report any mediator. On such occasions he reports "working on the pair" with a "it just came to me" type of answer. It may be that such unassignable associations are of an unconscious nature, as suggested by Bugelski and Scharlock (1952), Russell and Storms (1955), and Ryan (1960).

The vast amount of research done on lists of pairs and on serial lists must be viewed from the limited viewpoint of how *lists* are learned, and not how pairs or other units are learned. From the viewpoint of a learning theory, it is the latter question that is of greater importance, however practical the former may be.

While the present study was not directed at the question of massed as opposed to spaced learning it does suggest that such studies too might take into consideration the question of total time involved in the learning sessions. Typically such studies provide some rest interval between trials for the spaced group while the massed *S*s carry on with the activity. Commonly enough it is found that the spaced group performs at a higher level than the massed group after the same number of trials. When the spaced group's rests are included, however, it might be found that the total time was far in excess of the apparent advantage in trials, and the massed *S*s have learned proportionally more than the spaced.

The present finding must be interpreted as restricted to the experimental conditions involved, particularly with respect to the time intervals employed, the method of exposure, and the nonsense material task. This report must be considered more suggestive than conclusive and other tasks must be examined before greater confidence can be placed in the constant time principle.[3]

SUMMARY

Five groups of 20 *S*s learned eight pairs of nonsense syllables under conditions where the stimulus syllable was always presented for 2 sec. while the response syllable was presented for an additional 2, 4, 6, 8, or 15 sec. depending on the group. A 2-sec. interval preceded each new stimulus. The total exposure times for the several groups were 6, 8, 10, 12, or 19 sec. per syllable pair. A significant difference was found between trials to learn with the fastest learning occurring with the longest presentation times. When presentation time was multiplied by trials, however, no significant differences were found. It was concluded that T_p (presentation time) $\times T$ (trials) is a constant value.

Interrogation of *S*s after learning revealed that in 67% of the total possible learning units, *S*s made use of mediational devices. Those pairs that lent themselves most readily to translation into two meaningful words or could be combined into one word, or could initiate some imagery were most easily learned. These findings are discussed in terms of their significance for studies of "one-trial" learning.

REFERENCES

BUGELSKI, B. R., & SCHARLOCK, D. An experimental demonstration of unconscious mediation. *J. exp. Psychol.*, 1952, **44**, 334–338.

HOVLAND, C. I. Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.

MURDOCK, B. B. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

ROCK, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

RYAN, J. J. Comparison of verbal response transfer mediated by meaningfully similar and associated stimuli. *J. exp. Psychol.*, 1960, **60**, 408–415.

(Received February 21, 1961)

---

[3] It has been suggested to the writer by the linguist, Henry Lee Smith, that there is a striking similarity in the ages at which children of different nationalities learn to speak their native language to about the same level of efficiency.

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 413–414

SUPPLEMENTARY REPORT: GENERALIZATION OF A NONVERBAL RESPONSE TO AGGRESSIVE VERBAL STIMULI[1]

JAMES H. GEER AND ARNOLD H. BUSS

University of Pittsburgh

Buss (1961) investigated the effects of direction of generalization and intensity of vocal response on stimulus generalization with aggressive verbal stimuli. Generalizing from intense to mild aggressive stimuli yielded steeper gradients than generalizing from mild to intense stimuli. There was also a tendency to match the intensity of vocal response to the intensity of aggressive stimuli. The present study sought to test the generality of these findings with a nonverbal response and with an amplitude measure of response strength.

Method.—The response was a pull on the handle of an automobile parking brake mounted near *S*'s left knee at a 45° angle from the floor. The brake remained at its maximum excursion after each pull until deliberately returned to the starting point. The pause at the end of handle travel allowed *E* to read a pointer position on a scale mounted on the side of the apparatus, indicating the amount of handle travel. This scale value was converted to pounds of pull, the measure of response strength. Directly in front of *S* was a memory drum which presented the stimuli every 8 sec.

The *S*s were pretrained to make "hard" and "soft" pulls on the brake. A hard pull was ≥ 30 lb. and a soft pull was ≤ 21 lb. All *S*s were easily able to learn this discrimination and to make either response correctly upon request. The *S*s were told to make only hard or soft pulls throughout the experiment.

Two groups of *S*s generalized from intensely hostile stimuli, and their training series consisted of the 20 most hostile words. Group I was told "Right" for hard pulls and "Wrong" for soft pulls; Group III was told Right for soft pulls and Wrong for hard pulls. The other two groups generalized from mildly hostile stimuli, and their training series consisted of the 20 least hostile words. Group II was told Right for hard pulls and Wrong for soft pulls; Group IV was told Right for soft pulls and Wrong for hard pulls.

At the end of the 20 training trials *S*s were instructed to continue responding, but *E* would no longer tell them whether they were right or wrong. Then the generalization series was begun. This series consisted of 126 words varying randomly from mild to intense hostility, as scaled by college *S*s (Buss, 1961). The *S*s were 60 undergraduate males enrolled in introductory psychology classes.

Results.—The *S*s trained on high hostile words saw the word BLOODTHIRSTY first; the *S*s trained on low hostile words saw the word DEBATE first. The amplitude of response to the first word prior to any reinforcement was categorized as "Strong" (>25 lb.) or "Weak" (<25 lb.) Twenty-four of the 30 *S*s pulled strong to the word BLOODTHIRSTY and 21 of the 30 pulled weak to DEBATE. A chi square analysis of this distribution was significant beyond the .005 level. Thus there was a strong tendency to match response intensity with stimulus intensity prior to training.

Figure 1 shows the generalization gradients for the groups. The upper two gradients represent the hard pull groups; for these groups amplitude of response should diminish with distance from the training

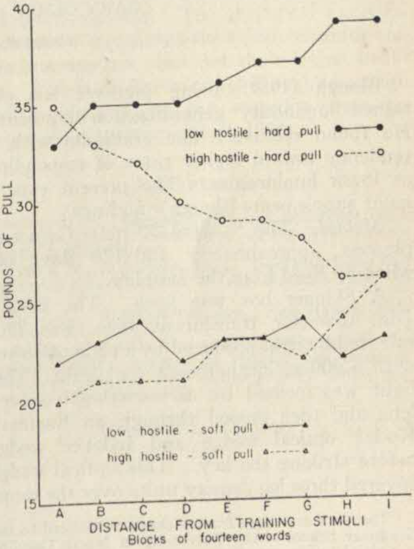


FIG. 1. Generalization gradients of response amplitude.

---

[1] Adapted from a MS thesis by the first author under the direction of the second author. The research was supported in part by Research Grant M-2681 from the National Institute of Mental Health, United States Public Health Service.

stimuli. The gradient of Group I does diminish, but Group II shows an inverse gradient: amplitude of response increases with distance from the training stimuli. An analysis of variance yielded a Groups × Stimuli F (the statistic relevant to slope differences) of 28.25 for 8 and 224 df ($P = .001$). Slopes were tested individually, and both were found to be significant ($P = .001$).

The lower two gradients represent the soft pull groups. Because the response measure is amplitude rather than frequency, the stimulus generalization gradients for these groups should rise, i.e., as the stimuli become more distant from the training stimuli the response should become less soft and therefore of greater amplitude. Group III's gradient does rise, but Group IV has a flat gradient. An analysis of variance yielded a Groups × Stimuli F of 4.21 for 8 and 224 df ($P = .001$). Slopes were tested individually; Group III's slope is significant, but Group IV's slope is not.

Discussion.—There was a tendency to match the intensity of the response with the intensity of the hostile stimuli. This was shown by both *Ss*' responses to the first training word and the inverse gradient of Group II. Direction of generalization was also found to be a significant variable.

These two determinants, matching and direction of generalization, were also found to be important in Buss' (1961) study. He found that matching was a more potent determiner of the slope of generalization gradients than direction of generalization. In the present study direction of generalization was the more potent determiner: Groups I and III had steep gradients, whereas Groups II and IV had flat or inverse gradients. The reason for the discrepancy may be the nature of the response. It seems likely that the tendency to shout hostile words would be stronger than the tendency to pull harder on a brake handle to hostile words. Thus a verbal response, such as Buss used, would probably lead to closer matching of response intensity than a nonverbal response.

REFERENCE

Buss, A. H. Stimulus generalization and aggressive verbal stimuli. *J. exp. Psychol.*, 1961, **61**, 469–473.

(Received March 31, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 414–415

SUPPLEMENTARY REPORT: STIMULUS GENERALIZATION GRADIENTS ALONG A LUMINOSITY CONTINUUM[1]

GARY OLSON AND RICHARD A. KING

University of North Carolina

Blough (1959), using pigeons, has obtained luminosity generalization gradients. He found relatively flat gradients with a tendency toward higher rates of responding at lower luminosities. The present experiment supplements Blough's findings.

Method.—The *Ss* were 20 male Carneaux pigeons, approximately 100–120 days old when received from the supplier.

A Skinner box was used. The key, a ⅞-in. diameter translucent disc, was illuminated by light produced by a slide projector with a 300-w. incandescent (ac) bulb. This light was focused by an additional convex lens and then passed through an Eastman Kodak optical wedge and balance wedge before striking the key. This optical wedge covered three log density units over the range of 15 setting positions. The settings used in this experiment were 2, 4, 6, 8 (CS), 10, 12, and 14 on the optical wedge. Each setting represents an increase of .4 log density units over the next smaller setting. Absolute luminosity values, obtained with a Macbeth illuminometer, were: Setting 2–13.89 mL.; Setting 8 (CS)-1.29 mL.; and Setting 14-0.11 mL. This is a much smaller range than that used by Blough. Blough also used a red colored disc.

The *Ss* were key-peck trained after reaching 80% of their ad lib. weight. They were then put on a 1-min. variable interval (VI) schedule. Every minute of this training was divided into two periods: (*a*) a 50-sec. responding period with the key illuminated at Setting 8 (the CS); and (*b*) a 10-sec. blackout period in which there was no possibility of a reinforcement and the key was darkened by a shutter. Each daily session was 30 min. long.

---

[1] The data are taken from a thesis submitted to the Graduate School of the University of North Carolina in partial fulfillment of the requirements for the MA degree, Olson (1960).

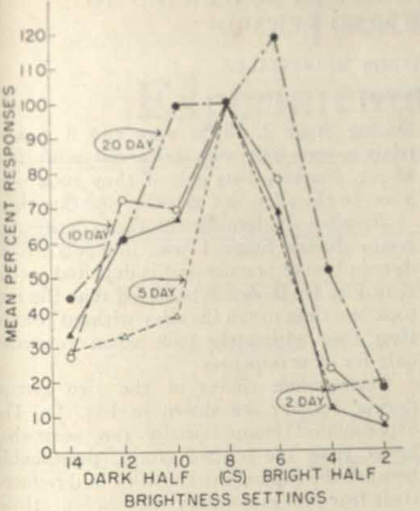


FIG. 1. Relative generalization gradients for the four training groups.

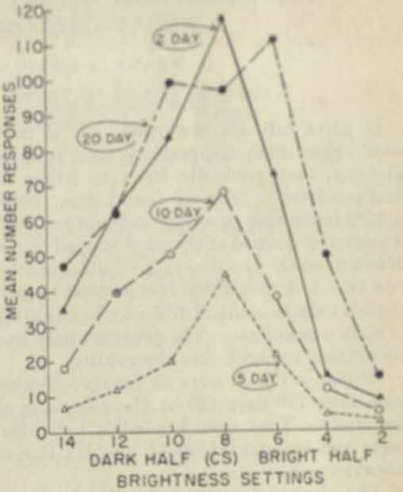


FIG. 2. Absolute generalization gradients for the four training groups.

The *S*s were taken directly from the lighted cage room to the darkened box, with no time allowed for dark adaptation. Since a second concern of this experiment was the effect of the amount of training on gradient slope, four groups of 5 *S*s each were given 20, 10, 5, or 2 days of VI training. The 2-day group consisted of *S*s specially selected for high response rate.

The generalization gradients were obtained during 2 consecutive days of extinction which began the day after the last training session. On Day 1 of extinction, there were 10 series of stimuli, and on Day 2, 5 series. Each stimulus series consisted of the seven light intensity values, ordered randomly. This is essentially the method of Guttman and Kalish (1956). The 50-sec. responding period, 10-sec. blackout cycle was maintained. During each 50-sec. period, *S* was given a different key illumination value. The wedge setting was changed during the 10-sec. blackout. The number of responses during each responding period was recorded, thus giving a total of 15 scores for each of the seven stimulus values.

Results.—The absolute and relative gradients for the four groups are shown in Fig. 1 and 2. The absolute measure is simply the total number of responses to each of the seven stimulus values. The relative measure is a percentage computed as follows:

$$\frac{\text{Total number of responses to a given stimulus value}}{\text{Total number of responses to the CS}} \times 100$$

The shapes of the gradients, like Blough's, are relatively flat. The statistical analyses indicated a significant dark preference for the absolute measure, but not the relative measure. This is similar to one of Blough's findings and is opposed to the stimulus intensity dynamism construct (Hull, 1952).

REFERENCES

BLOUGH, D. S. Generalization and preference on a stimulus-intensity continuum. *J. exp. Anal. Behav.*, 1959, **2**, 307–315.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

OLSON, G. A study of two variables affecting stimulus generalization: Amount of training and stimulus intensity. Unpublished master's thesis, University of North Carolina, 1960.

(Received April 3, 1961)

Journal of Experimental Psychology
1962, Vol 63, No. 4, 416

SUPPLEMENTARY REPORT: SHIFT FROM NONREWARD TO NEGATIVELY CORRELATED REWARD[1]

FRANK A. LOGAN AND LOUIS M. GONZALEZ

Yale University

If naive rats are rewarded only if they take longer than a specified time running an alley, they gradually learn to adjust to this condition. Trapold (see Logan, 1960, p. 190) found that rats that had first run with reward uncorrelated with speed learned to run slowly as easily as naive rats. Subsequently, rats that had previously run without reward seemed to have unusual difficulty in adjusting to such a condition. The present study was undertaken to check this observation.

Method.—The *S*s were 24 Sprague-Dawley male rats, 120 days old at the beginning of the study. They were housed in individual cages and maintained on 12 gm/day Purina laboratory chow.

The double-alley maze and procedures used in this study have been described elsewhere (Logan, 1960). During Stage 1, each *S* in the experimental group was placed in a delay box, then run 1 or 4 trials in a 4-ft. straight alley without reward, and then placed in a second delay box before being returned to its home cage. This training lasted for 76 trials during which time *S*s in the control group were shuttled through the delay boxes each day but were not exposed to the alley. During Stage 2, all *S*s were run 6 massed trials a day and were rewarded with four 45-mg. Noyes pellets only if they took over 3 sec. in the alley, not counting starting time.

Results.—Performance by the experimental group during Stage 1 was independent of distribution of practice and is depicted by the X in Fig. 1. It should be noted that the rats took less time to run the alley without reward than they ultimately took when rewarded only for slow responses.

Performance curves of the two groups during Stage 2 are shown in Fig. 1. The experimental group initially ran somewhat faster than the control group, presumably because the exposure to the alley had reduced their fear and exploratory tendencies. However, the curves are closely comparable through most of their course and compare favorably with those obtained from naive *S*s. Hence, it is concluded that experience of nonreward does not affect adjustment to conditions of negatively correlated reward.

REFERENCE

LOGAN, F. A. *Incentive: How the conditions of reinforcement affect the performance of rats.* New Haven: Yale Univer. Press, 1960.

(Received April 22, 1961)

---

[1] Research supported in part by the National Science Foundation.

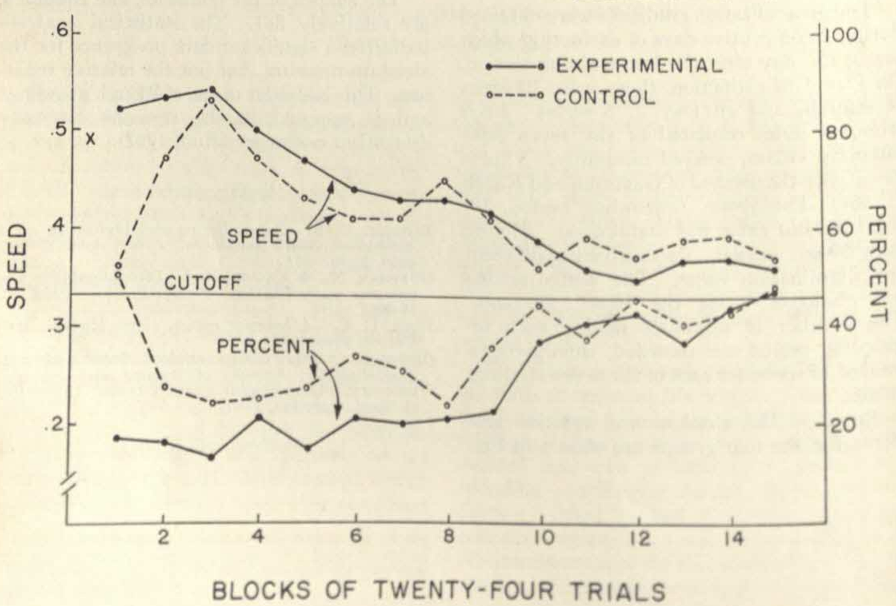


FIG. 1. Running speed and percentage below cutoff of *S*s that were rewarded only if they took over 3 sec. in the alley. (The experimental group had had prior exposure to the alley without reward, their performance at the end of this stage being indicated by the X; the control group had been handled but not previously exposed to the alley.)

Journal of
Experimental Psychology

Vol. 63, No. 5 May 1962

SET, SUGGESTION, AND CONDITIONING[1]

WILLIAM W. GRINGS, SIDNEY CARLIN
University of Southern California
AND MORTIMER H. APPLEY
Southern Illinois University

The concept of reinforcement, as it applies to classical conditioning of autonomic responses with human *S*s, has several different interpretations. To many workers the reference of the term is strictly empirical, whereas to other workers there is strong implication that the essential requirement is a change in a motivational state (Mowrer, 1960; Spence, 1956, 1960).

Some theorists have emphasized reinforcement as a *reactive* process within *S*. Woodworth (1947, 1958), for example, asserts that during conditioning *S* is learning to anticipate stimulation. Conditioning involves essentially "a change in his (*S*'s) way of receiving, or perceiving, the sequence of stimuli and especially the preliminary or conditioned stimulus" (Woodworth, 1947, p. 122).

Variations on the notions of reinforcement as some form of confirmation of an anticipatory reaction by *S* to the CS are numerous. Quite different discussions are provided by Hilgard and Marquis (1940), Bruner (1951), Anokhin (1955), and Solley and Murphy (1960). One common weakness of most "reinforcement of perception" concepts is that they do not have adequate operational components. The integration of concepts of perception and reinforcement was the subject of a paper by Razran (1955). And one of the present authors (Grings, 1960) has proposed operations for inferring a preparatory set concept in the context of classical conditioning.

The present experiment was set up to explore some of the implications of a reinforcement concept based on an assumed anticipation of stimulation. It was an attempt to demonstrate, with human *S*s, that a tendency to anticipate stimulation (i.e., a response set) can be manipulated in a situation analogous to classical conditioning. The operations for establishing the set were suggested by Cornbecker, Welch, and Fisichelli (1949) in a study of conditioning factors underlying hypnosis.

The conditioned response (CR) was defined as the GSR to a two-word verbal cue (CS) capable of suggesting stimulation. Two test cues, ELECTRIC SHOCK and SCALDING WATER, were considered initially (intrinsically) adequate to elicit GSRs, and a control cue PLEASANT DAY was used for reference purposes. The two test cues differed, however, with respect to their adequacy to suggest stimulation in the particular experimental setting. This difference was labeled cue *feasibility*. ELECTRIC SHOCK was called feasible since an *S* with GSR and ground electrodes attached might reasonably expect an actual shock. On the other hand, the cue SCALDING WATER was called unfeasible since the otherwise barren experimental cubicle contained no apparent means for presenting the actual stimulus. (The neutral cue PLEASANT DAY was assumed not to be suggestive of actual stimulation.)

---

[1] This research was supported by a grant (G-8612) from the National Science Foundation.

Changes in GSR to the three test cues were studied as a function of two independent variables: feasibility, as defined above; and amount of reinforcement. Reinforcement was not employed in the usual sense of the term, for the reinforcing state of affairs (presentation of a UCS) was not applied directly to the test cues previously mentioned. That is, the cue ELECTRIC SHOCK was never actually followed by a shock, or SCALDING WATER followed by scalding water. Instead, seven other cues were presented and were followed by appropriate stimulation.

The single sequence of one verbal cue followed by the stimulation it suggested was arbitrarily labeled as one reinforcement. As experimental variation two groups were used: a *long* training group received seven series of the seven cues plus a test series; the other, a *short* training group, received two series of the training cues plus a test series.

METHOD

Subjects.—Eighty paid volunteers, 49 men and 31 women, from undergraduate psychology classes at the University of Southern California were observed for periods of from 45 min. to 90 min.

Apparatus.—The GSR was obtained as a dc resistance through a zinc, saline paste, skin junction employing $\frac{3}{4}$-in. diameter Lucite cup electrodes fastened by rubber straps to the palm and back of *S*'s nonpreferred hand. The electrodes led to a modified Darrow-type bridge the output of which was amplified by a Brush, Model BL 530 dc amplifier and recorded on an Esterline-Angus ink writing milliameter. Current through *S* was 45 μa.

The *S*s were seated at a table in one section of the laboratory. By appropriate switching in another portion of the laboratory *E* could cause (*a*) an 8-in. fan to blow across *S*, (*b*) a green flood light to come on, (*c*) tape recorded symphonic music to play, (*d*) all room lights to go out, (*e*) a 100-db., 1000°-cps tone to sound, (*f*) a light bulb in front of *S* to blink at the rate of 10/sec, or (*g*) a mechanical transducer on *S*'s finger to vibrate. The *E* could also project verbal cues from a 2 × 2 in. projector onto an 8 × 10 in. ground glass screen in front of *S*. The verbal cues corresponding to the forementioned stimulations were respectively (*a*) COOL BREEZE, (*b*) GREEN LIGHT, (*c*) SOFT MUSIC, (*d*) DARK ROOM, (*e*) LOUD TONE, (*f*) FLICKERING LIGHT, and (*g*) FINGER VIBRATION.

Verbal cues were presented for 10 sec. per trial. During the second half of this period the appropriate UCS was presented. Timing of this sequence was controlled by electronic interval timers. Intertrial intervals were varied randomly among 25, 30, and 35 sec.

General design.—As originally conceived the study was to employ four different treatment groups of 10 *S*s each in a 2 × 2 factorial arrangement with feasible vs. unfeasible test stimuli on one axis and long vs. short training conditions on the other axis. The first 40 *S*s were tested in this arrangement. However, a second 40 *S*s were tested, duplicating this condition, with the exception that instead of stopping the reinforcement of the *short* group after it was tested following 14 reinforcements, this group was given 35 additional reinforcements and tested again at the *long* test point. This made it possible to use 30 different *S*s for evaluation of certain cell means.

General procedure.—The *S*s were told "This is a study of your palmar skin resistance (the sweat gland activity of the palms of your hands)." Then after a brief discussion of the GSR they were instructed to "Watch the screen and read aloud the words that will appear on it from time to time."

The *S*s were first assigned by a table of random numbers to one of the four experimental groups: (*a*) short training series, feasible test cue; (*b*) short series, unfeasible test cue; (*c*) long series, feasible test cue; and (*d*) long series, unfeasible test cue. Then all were presented a pretest series which included all subsequently used cues and 10 additional "filler" cue phrases.

The short reinforcement condition contained two series of the seven reinforced cues plus a test series which contained the sequence of reinforced cues with test cues introduced at Ordinal Positions 1 and 6 with the series ending in a disparity trial, i.e., the verbal cue GREEN LIGHT being followed by the actual loud tone. The disparity trial was omitted from the short group during testing of the last 40 *S*s since these *S*s received further training and test following 35 additional training trials. For any particular *S* the two test cues would be either the feasible or unfeasible cue and the reference cue. The order of presentation (between Positions 1 and 6) was varied systematically.

The long reinforcement condition contained seven series of the seven reinforced cues plus the test series, as described above. Here, too, the test series ended in a disparity trial.

Following the test series, *S*s were given a questionnaire to fill out. It posed such questions as "Did you expect to feel scalding water?" "Did you experience scalding water?" "Did you expect an electric shock?" "Did you experience an electric shock?"

RESULTS AND DISCUSSION

Table 1 shows the means and standard deviations of the magnitudes of response to the three critical or test cues as a function of the position in the reinforcement series. The basic unit of response is the square root of the maximum change in conductance, in micromhos, occurring during the 10 sec. following presentation of the cue.

TABLE 1

GSRs (MICROMHOS) TO TEST CUES AT DIFFERENT LEVELS OF TRAINING

Cues	Training Level		
	Pre-training	14 Reinforcements	49 Reinforcements
ELECTRIC SHOCK			
Mean	1.03	1.34	1.14
SD	.69	.72	.67
N	80	30	30
SCALDING WATER			
Mean	.79	.94	.64
SD	.59	.61	.53
N	80	30	30
PLEASANT DAY			
Mean	.73	.64	.59
SD	.55	.53	.53
N	80	60	60

A few review comments may help the interpretation of the data in this table. All *S*s were tested at each initial point, hence the *N*s for comparisons of pretest response magnitudes are 80. When pretest comparisons are made, the response to suggested shock is significantly greater than the response to either of the other stimuli. Because of this initial difference among stimuli, it is the *change* in magnitude of response to these cues that is emphasized.[2] When changes are evaluated the *N*s at each point equal 30. These samples are composed of 10 *S*s from the first replication of the study and 20 *S*s from the second replication at which time *S*s were tested at both the long and short training points.[3]

---

[2] In subsequent experimentation with these same stimuli, the initial differences among stimuli have been removed by administration of an adaptation series at the outset of the experiment.

[3] A comparison of the responses at the late training point, for *S*s who had the interposed test at the early training point and *S*s who had none, showed no difference. Data from the last "replication" of the experiment are therefore used at both the early and late training test points. Measures of change

TABLE 2

MEAN ADJUSTED SCORES FOR TEST STIMULI DURING PRETRAINING AND POST-ACQUISITION TEST PERIODS

Cue	Short Reinforcement Series		Long Reinforcement Series	
	Pre-training	Post-test	Pre-training	Post-test
Feasible	.15	.75	.37	.58
Unfeasible	.01	.20	.02	.22

Note.—$N = 80$; 20 *Ss* in each subgroup.
Difference in magnitude of GSR to individual test and reference stimuli measured in micromhos, square root of conductance change.

When changes are evaluated as the mean difference from pretraining to posttraining response levels, the change for the feasible cue is the only one which is statistically significant ($t = 2.76$; $df = 29$). That is, the early training experience increases the response evocation capability of the feasible cue. There also appears to be an early training increase for the unfeasible cue, but it is not a statistically significant one. Somewhere after this first test point the response evocation falls, for when tested at the late training point neither stimulus is significantly different from the pretraining value. It will be noted that the neutral or reference stimulus shows a consistent course of adaptation throughout the experiment. It is therefore quite probable that the training series was so long that general response adaptation occurred, competing strongly with the learning trends presumably operating early in training.

To confirm that there was adaptation, responses to the reinforced cues (LOUD TONE, COOL BREEZE, etc.) were plotted for successive trials. A typical curve of decreasing magnitude of response with repeated trials was obtained.

One way of partialing out general and situational adaptation is to subtract from the GSR elicited by the crucial test stimulus the GSR given to the neutral or reference stimulus on the closest adjacent trial. This is based on the assumption that the response to the test stimulus represented effects of both learning and adaptation whereas the response to the reference stimulus reflected only general situational adaptation. Tabulations and analyses were made using this adjusted score (the difference between response to test and reference stimuli) as the dependent variable. Mean values for the major experimental categories are given in Table 2.

To evaluate the significance of the differences evident in this table an analysis of variance as a mixed factorial design was performed. Feasibility and reinforcement are between-subject main effects and pretest vs. posttest periods provide a within-subject main effect. Three components are significant beyond the 1% level, as follows. The overall feasibility variable is significant ($F = 7.89$; $df = 1/76$). Here, it is important to recall that this test reflects pretest as well as posttest differences between feasible and unfeasible stimuli.

The overall change from pretest to posttest conditions is also significant ($F = 12.77$; $df = 1/76$); whereas, the difference between the two posttest groups (long reinforcement vs. short reinforcement) is not significant. The triple interaction of feasibility, reinforcement groups, and pre-to-posttest periods is also significant ($F = 14.22$; $df = 1/76$).

In summary, the analysis of variance on difference scores confirms the analysis of response means of Table 1. The principal variation

---

due to training were made on mean differences of samples of 30 *S*s for each point on the curve.

occurs for the feasible stimulus cue during the short series of reinforcements.

To further evaluate the extent of general situational adaptation a second form of response measure, the perceptual disparity response was used. This involves presenting a CS followed by a UCS which differs in some specified way from the UCS with which the CS had been paired previously. In this case the cue GREEN LIGHT was presented, then at the end of 5 sec. the actual loud tone was given. The disparity response was computed as the difference in response to the actual tone UCS when it was preceded by the wrong or disparate cue GREEN LIGHT and the last previous response to the tone UCS when it was preceded by the cue LOUD TONE. The fact that this perceptual disparity response (for the 40 *S*s on whom it was available) was significantly larger for the long reinforcement groups than for the short reinforcement groups was taken as further evidence that a significant degree of adaptation had occurred between 14 and 49 reinforcements. The data on this comparison have been reported elsewhere (Grings, 1960).

It had been predicted at the start of the study that the tendency to respond to an unfeasible cue would show up more strongly with increased training. In other words, the longer the reinforcement series, the stronger the set might be to expect stimulation on the scalding water cue. This prediction was not upheld, for response tendencies to both cues appeared to adapt with long training series; and this result was the same whether or not *S*s had been tested at the short training point.

Some verification of the definition of the feasibility dimension can be obtained from a tabulation of data from the postexperimental questionnaire. Of the 80 *S*s, 67 reported that they expected to receive an electric shock following the verbal cue; whereas only 12 said that they expected to experience scalding water. When asked whether they actually experienced an electric shock, 12 said yes; whereas 1 of the 80 reported experiencing the scalding water.

The evidence supports the notion that perceptual sets in the form of tendencies to anticipate stimulation can be manipulated experimentally. By presenting verbal test cues during a period when other verbal cues are being reinforced, the capability of the test cues to elicit anticipatory GSRs is enhanced. The changes that occur in the response tendency (set) would appear to be first one of increase with associated reinforcement training, then a reduction (adaptation effect) accompanying further training.

It is proposed that the experimental situation discussed has usefulness in studies of the reinforcement process with human *S*s, particularly where concern is for modification of sets to anticipate stimulation. The situation does not involve direct reinforcement but employs a kind of atmosphere effect wherein nondirect reinforcement can be influential in modifying a response set. It can be further emphasized that such nondirect reinforcement may be the only type that is effective in such a situation. If the noxious suggestion cues were reinforced directly, it is probable that response adaptation would begin immediately; whereas, by focusing on the anticipation process and confirming or reinforcing some anticipations, a phenomenon is observed which would not otherwise occur.

SUMMARY

A set or tendency to anticipate that stimulation will follow a verbal cue suggesting such stimulation is inferred from the magnitude of GSR to the cue. The experimental procedures for modifying the set are analogous to classical conditioning procedures, except that the reinforcing state of affairs is not applied directly to the test cues but is applied instead to a series of seven different cues.

Experimental variation is introduced in a 2 × 2 factorial arrangement with one inde-

pendent variable the number of reinforcements applied before test, and the second independent variable the nature of the test cue. Half of the 80 *S*s were tested on a cue involving a feasible suggestion of stimulation. The other half were tested on a cue designated unfeasible.

Plots of magnitude of response to the critical test cues as a function of position in the reinforcement series showed an increase of response to both cues at the short training test point, then a decrease with further training. The principal change occurred for the feasible stimulus during the short series of reinforcements.

REFERENCES

ANOKHIN, P. K. Characteristics of the afferent apparatus of a conditioned reflex and its importance for psychology. *Vop. Psikhol.*, 1955, No. 6, 16–38.

BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951.

CORNBECKER, F., WELCH, L., & FISICHELLI, V. Conditioning factors underlying hypnosis. *J. abnorm. soc. Psychol.*, 1949, **44**, 212–222.

GRINGS, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **67**, 243–252.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

RAZRAN, G. H. S. Conditioning and perception. *Psychol. Rev.*, 1955, **62**, 83–95.

SOLLEY, C. M., & MURPHY, G. *Development of the perceptual world.* New York: Basic Books, 1960.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WOODWORTH, R. S. Re-enforcement of perception. *Amer. J. Psychol.*, 1947, **60**, 119–124.

WOODWORTH, R. S. *Dynamics of behavior.* New York: Holt, 1958.

(Received February 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 423–425

ACUITY AND THE STATISTICAL THEORY OF FIGURAL AFTEREFFECTS

F. H. GEORGE

University of Bristol

An experiment was carried out on visual acuity under conditions of fatigue similar to those met in the figural aftereffect situation. This experiment was performed with the object of testing a supposition made to preserve the general form of the Osgood-Heyer (1952) type of explanation against some objections made by Deutsch (1956). The supposition was that whereas in the ordinary course of events summation of excitation, to the extent suggested by Osgood and Heyer, does *not* occur in the visual cortex, under conditions of fatigue, such as in the inspection period of the figural aftereffect experiments, there is a spread of excitation that brings about summation over very considerable areas. This effect was thought to be more marked with respect to areas in the periphery of the retina. A simple way of testing this supposition offers itself in the use of a situation that demands discrimination of two contour lines under "fatiguing" conditions. The results obtained suggest possible modification to the Osgood-Heyer model.

The principal fault in the Osgood-Heyer (1952) model, according to Deutsch (1956), is that it uses an altogether too versatile concept of summation. The apparent shift of the contour line in the figural aftereffect situation depends upon the new contour line falling on the area already fatigued, the distributions thus overlapping and causing summation. But the argument goes on to say that these adjacent lines are easily discriminated within the area over which the figural aftereffect shift is supposed to take place. Now clearly Osgood and Heyer cannot have it both ways; the problem exists as to how we can discriminate lines at all, *within* the range of the shift, for either there is a failure of discrimination—where, incidentally, it is unquestionably known to be possible—because overlapping distributions summate with a resulting shift, or we can discriminate, in which case we do not get a shift.

METHOD

Subjects.—Twenty *S*s were used; they were all ordinary citizens of Bristol or undergraduates of the university. No prior check was made on their eyesight.

Apparatus.—The apparatus was made up of rectangular pieces of 8 × 12 in. white cardboard, on each of which were drawn pairs of vertical lines and fixation points.

Procedure.—The cards were presented to *S*s in a small stand and under standard conditions of illumination. There is little doubt that the results would be affected to some extent by the strength of the illumination, but this experiment was designed merely to discover whether the ordinary discrimination of lines would occur when slightly displaced from the fovea, and under conditions of "fatigue."

Three different pairs of adjacent lines were used at three different distances from a fixation point; there was one pair of lines on each card, each line being 2 in. long and $\frac{1}{16}$ in. thick. On the first card (A) the lines were $\frac{1}{16}$ in. apart, and the three fixation points were 2, 4, and 6 in., respectively, from the nearest of the pairs of lines. On Card B the two lines were $\frac{1}{8}$ in. apart, and the fixation points were the same distance away from the nearest line as on Card A. On Card C the lines were $\frac{3}{16}$ in. apart, and again the fixation points were the same distances from the nearest line as they were on Card A. The cards were

TABLE 1

NUMBER OF *S*s SEEING ONE LINE, TWO LINES, OR NEITHER (INDEFINITE)

Card	Fixation Point	Binocular Viewing						Monocular Viewing					
		Lines to Right			Lines to Left			Lines to Right			Lines to Left		
		1 Line	2 Lines	Indef.	1 Line	2 Lines	Indef.	1 Line	2 Lines	Indef.	1 Line	2 Lines	Indef.
A	2 in.	9	0	1	10	0	0	8	0	2	9	0	1
	4 in.	8	1	1	8	1	1	8	1	1	10	0	0
	6 in.	9	0	1	8	0	2	9	0	1	7	2	1
B	2 in.	9	0	1	8	1	1	7	2	1	9	1	0
	4 in.	8	1	1	7	2	1	8	1	1	8	0	2
	6 in.	8	2	0	7	1	2	8	1	1	7	2	1
C	2 in.	4	3	3	2	7	1	1	7	2	2	7	1
	4 in.	1	5	4	0	8	2	0	10	0	0	9	1
	6 in.	0	8	2	0	7	3	0	9	1	0	7	3

all viewed from a distance of 18 in. from the midpoint between the eyes.

The *S*s were instructed to fixate a particular fixation point, and to continue to fixate it for 1 min., at the end of which *E* said, "Keep fixating the fixation point carefully, and tell me how many lines you are aware of on the left, now." The word "right" was used in place of "left" when the lines were on the other side.

This same request was repeated after a further 30 sec. of fixating. The lines on the card were not made visible until some seconds after *S* had started fixating, so that he did not have an ordinary, clear, prior view of the pairs of lines.

The *S*s were asked to view all three fixation points for each card, first with the lines on the left and then with the lines on the right. They were then asked to repeat the whole procedure monocularly, the verbal instructions being amended to read, "Close one eye and fixate the fixation point with the other eye." After 1 min. the instruction continued, "Now open the closed eye and shut the other one, and fixate the same fixation point, and tell me how many lines there are on the left (or right) of the fixation point."

As a precaution, the order of presentation was sometimes changed to monocular-binocular, as opposed to the above order, and within this the nine different cases were presented in a random order.

RESULTS AND DISCUSSION

Since the experiment was not really intended to be a highly refined one, but merely a test of an hypothesis of a straightforward kind—although requiring controlled conditions—the results were scored simply as one line or two or indefinite, according to *S*'s reply when questioned. Table 1 shows all the results in tabular form. It should be added that a control group of 20 *S*s was used, and for all positions of the lines they were able to detect the presence of two lines when asked within 5 sec. of viewing the figures. This control group did not, of course, go through any fixation period (inspection period).

The results in Table 1 are clear cut in that a significantly large number of *S*s under all the conditions saw the two lines disappear into one for all of cases of Card A and all of cases of Card B, but only a few for all of cases of Card C. These figures are, of course, highly significant. The rows marked "indefinite" simply contain the results of those *S*s who were unable to say whether they saw one or two lines. But first it must be remarked that many *S*s recorded many other observations of interest, although for the most part familiar to workers in the field of the figural aftereffect, and these cannot be discussed here.

The results offer confirmation of the argument that the effect of fatiguing the eye by steady fixation is to spread the distribution of excitation in such a way as to cause summation to occur much more readily in the visual cortex. The fact that it was nearly as effective monocularly, when the eyes were exchanged before the decision was taken, shows that the effect is almost certainly central.

It should be mentioned that many more *S*s were able to see the effect informally; in fact, anyone may draw two such lines, with a dot for fixation purposes, and convince himself of the facts.

This all means that the principal objection to the Osgood-Heyer model has been circumvented, and that it is perfectly possible to account for both figural aftereffects and ordinary visual discrimination, still within the ordinary interpretation of summation.

Nevertheless, some amendments do have to be made to the Osgood-Heyer model. They are to the effect that the spread of distribution is a function of the fatiguing, and not merely that with fatigue there is a diminution of excitation. Similarly, but for other reasons, they may have to drop the idea that eye movements are *necessary to* the original distribution of excitation (Deutsch, 1956; Riggs, Ratliff, Cornsweet, & Cornsweet, 1953). But there is now some additional, although perhaps minor, reason for believing this to be the case, and that is that we may expect a considerable disparity between the girth of the distribution of excitation in the ordinary course of a line being projected onto the retina, and the case where the line is projected onto the retina when fixation of some other point occurs for a period of 1 min. or more. The difference here could be between the narrow ridge of excitation caused by the line, and the spread of excitation caused by fatigue. No attempt will be made here to relate the assumed spread of excitation with fatigue to the change of activity in the retinal elements.

SUMMARY

This experiment was an attempt to reinstate the Osgood-Heyer (1952) model for explanatory purposes in perception, with the object of making plausible extension of that model in order to explain movement aftereffects.

The experiment involved sets of double-line figures at various distances from a fixation point and *S*s were unable to distinguish the two lines when sufficiently far from the fixation point, *and* under conditions of "fatigue." Thus, the principal features of the Osgood-Heyer model can easily be preserved from the main objections made against them by Deutsch (1956).

REFERENCES

DEUTSCH, J. A. The statistical theory of figural aftereffects and acuity. *Brit. J. Psychol.*, 1956, **47**, 208–215.

OSGOOD, C. E., & HEYER, A. W. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

RIGGS, L. A., RATLIFF, F., CORNSWEET, J. C., & CORNSWEET, T. N. Disappearance of a steadily fixated visual test object. *J. Opt. Soc. Amer.*, 1953, **43**, 495–501.

(Received January 31, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 426–437

ASSIMILATION VS. CONTRAST IN THE ANCHORING OF PERCEPTUAL JUDGMENTS OF WEIGHT[1]

ALLEN PARDUCCI AND LOUISE M. MARSHALL

University of California, Los Angeles

Successive perceptual contrast may be demonstrated for comparative judgments by interpolating an extra stimulus between the standard and comparison stimuli. The value of the comparison judged equal to the standard, the PSE, varies directly (and the judgments inversely) with the value of the interpolated stimulus (Guilford, 1954). Successive contrast can also be demonstrated in the absence of a specified standard, i.e., for absolute judgments. When a new stimulus value is added to a series presented for judgment by the method of single stimuli, the mean judgment of the remaining stimuli varies inversely with the value of the new stimulus. This contrast occurs both when the end-category of the absolute scale is anchored to the new stimulus (e.g., Rogers, 1941; Sherif, Taub, & Hovland, 1958) and also when the new stimulus is presented without specific anchoring instructions (Parducci, 1954; Postman & Miller, 1945).

These examples of perceptual contrast are quantitatively described by the theory of adaptation level (Helson, 1947, 1959). According to this theory, the PSE or the adaptation level, AL, the stimulus value at which the scale of judgment is centered, corresponds to a weighted logarithmic mean of the various physical values of the stimuli. Each of the judgments is in turn determined by the ratio of the stimulus being judged to AL. To the degree that the introduction of a new stimulus shifts the mean (and thus AL) toward its own physical value, contrast or an opposing shift in judgment occurs.

However, the results of a recent experiment with lifted weights have raised doubts concerning the generality of this description of judgment (Sherif et al., 1958). Although the expected contrast effects were found for anchors outside the range of the regular series, the opposite effect (i.e., heavier judgments of the series weights) was produced by an anchor equal in weight to the heaviest member of the regular series. Sherif et al. (1958) applied the term *assimilation* to this phenomenon, interpreting it as an analogue to certain propaganda effects of social communication. They also noted that such assimilation is not easily explained in terms of the adaptation-level approach.

The present experiments were designed to evaluate this apparently negative instance with respect to an alternative application of the theory of adaptation level. Instead of assuming that the anchoring stimulus serves as just another member of the series presented for absolute judgment, it is here proposed that the anchor should be treated as a standard for comparative judgment. Since the experimental instructions used by Sherif et al. (1958) state that the anchoring stimulus is to be judged with the heaviest category, PSE is by definition the physical value of the stimulus actually judged with the heaviest category. According to the theory of adaptation level, PSE is

---

[1] This research was supported in part by a grant (M-3134 A) from the National Institutes of Health.

equal to a weighted logarithmic mean of the standard and comparison stimuli (Helson, 1959; Michels & Helson, 1954). When the standard is identical to the heaviest comparison weight, PSE will be below the value of the standard since any weighted mean of the stimulus values would have to be lower than the standard. But if PSE is below the actual physical value of this anchoring stimulus, the heaviest comparison weight is perceived to be heavier than the anchor—perceptual contrast. The anchoring conditions employed by Sherif et al. (1958) thus lower the value of the stimulus to which the "heaviest" category of judgment is applied, and the entire scale of judgment is shifted downward.

The present hypothesis is that this downward shift, the assimilation found by Sherif et al. (1958), is a special case of the *central-tendency effect* (Hollingworth, 1910; Long, 1937). As described by the theory of adaptation level, the central-tendency effect is basically a phenomenon of perceptual contrast. The contrast lies in the shift in PSE toward a weighted mean of the experimental stimuli. The following experiments evaluate this interpretation by varying the instructions for judgment.

EXPERIMENT I: TEST FOR CENTRAL TENDENCY

The first objective was to replicate the relevant portions of the experiment reported by Sherif et al. (1958) with additional controls for the order of presentation of the respective experimental conditions. To test the present interpretation of assimilation, two additional conditions were introduced which permitted direct assessment of the PSE for the anchoring stimulus. For one condition, a different category of judgment was fixed to the anchoring stimulus; for the other, an additional category was added to the scale of judgment.

Method

Stimuli.—The six weights used as the "original stimulus series" by Sherif et al. (55, 75, 93, 109, 125, and 141 gm.) were also used for every condition of the present experiments. The individual weights were presented for judgment by means of an apparatus constructed in accordance with the description reported for the previous experiment. Essentially, it served to block *S*'s view of the weights and of *E*, the lifting being accomplished by pulling down a handle which was attached to a cord leading over a pair of low-friction wheels to the weight.

Procedure.—The instructions described by Sherif et al. (1958) were used for three of the present conditions. Thus for Cond. No-A (No-Anchor), *S*s had to make their judgments in terms of an absolute scale of 1 to 6, inclusive, with no special anchoring of any of the response categories (i.e., *S*s were not told the number which went with each weight, or even how many weights there were). For Cond. H-A (Heavy-Anchor), *S*s were told that the stimuli would be presented in pairs, the first member of each pair indicating the weight which they were to call 6. This was the 141-gm. weight. The 55-gm. weight was presented as the anchor for Cond. L-A (Light-Anchor), and *S*s were instructed to call it 1.

The 141-gm. weight was used as the anchor for both of the new conditions. It was again labeled 6 for Cond. 7-C (7-Category), but *S*s were instructed to make their judgments in terms of a 7- rather than a 6-point scale. The anchor was labeled 5 on a 6-point scale for Cond. Anchor-5. These conditions can perhaps be more readily distinguished by reference to Fig. 1, with special attention to the location of the different anchoring weights with respect to the response scale.

Each condition required a total of 288 judgments by each *S*. In addition, there were 288 unjudged presentations (the odd trials) for each of the anchor conditions. The order of presentation was randomized over blocks of 48 trials (or pairs of trials when an anchor was used) so that each of the six weights appeared eight times within each block. An experimental session lasted somewhat under 1 hr., with two short rests permitted within the session. Each *S* served in a

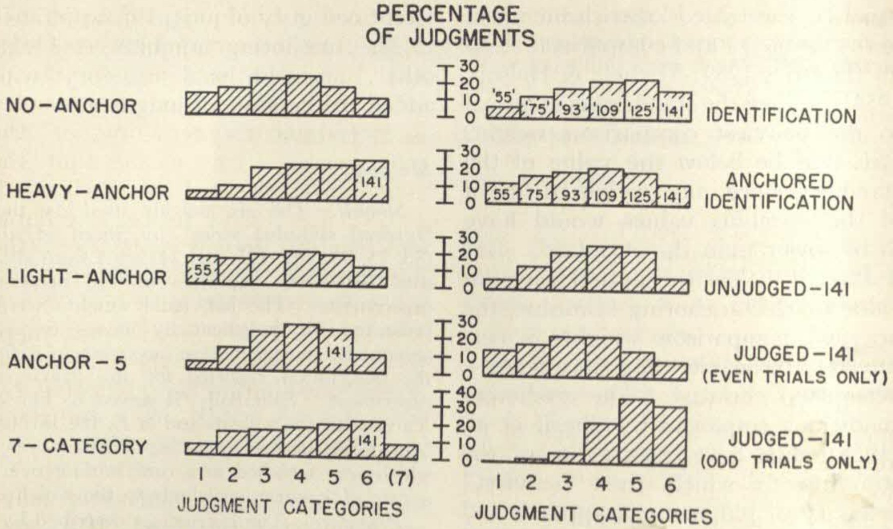


FIG. 1. Distribution of judgments for conditions selected from Exp. I, III, and IV. (Anchored categories indicated by weight, gm., of anchoring stimulus.)

minimum of two experimental sessions which were always on different days.

Subjects and order of the conditions.—Each of the groups serving in these experiments was composed of 10 *S*s drawn from the introductory psychology course at the University of California, Los Angeles. Following the procedure reported by Sherif et al. (1958), the two replication groups served under more than one of the conditions being compared.

TABLE 1

MEAN ADAPTATION LEVELS AND PERCENTAGE DISTRIBUTION OF JUDGMENTS

Condition	Categories of Judgment						AL	*SD*
	1	2	3	4	5	6		
No-Anchor (1st)[a]	13	16	21	23	17	10	103	5.8
Heavy-Anchor (1st)	5	9	20	21	21	24	87	11.4
No-Anchor (2nd)	6	13	21	24	20	16	94	7.1
Heavy-Anchor (2nd)	11	14	16	17	21	21	93	8.6
Anchor-5	5	13	22	27	23	10	94	6.7
Anchor-5 (3rd)	14	16	18	23	23	6	101	6.9
Light-Anchor	17	16	18	20	18	11	105	10.1
7-Category[b]	6	14	15	17	19	21	95	6.9
Heavy-Standard	15	24	33	22	5	1	124[c]	11.5
Light-Standard	0	2	11	31	33	23	56[c]	14.8
Unjudged-141	6	14	22	27	23	8	95	10.6
Judged-141	14	18	22	22	15	9	105	5.9
Judged-141[d]	7	9	14	22	27	21	114	8.3
Identification	7	13	18	22	23	17	93	9.2
Anch.-Identification	12	17	19	21	19	12	101	2.5

[a] Order of presentation indicated only when *S*s judged under more than one condition.
[b] Category 7 was used for 8% of the judgments.
[c] PSE for the standard.
[d] Including judgments of 141-gm. weight on odd trials.

However, the order of the conditions was systematically varied in the present design so that the effects of order could themselves be assessed. Thus one group was exposed first to Cond. No-A, then to Cond. H-A, and finally to a new condition, Anchor-5. A second group was exposed to the first two of these conditions only; for this group, Cond. H-A preceded Cond. No-A. A third group served only under Cond. Anchor-5. Each of the other groups in these experiments also served under only one of the experimental conditions.

Results and Discussion

The percentage of the 288 judgments made in terms of each of the respective categories was computed for each *S* under each condition. Table 1 shows these percentages averaged over 10 *S*s. To facilitate comparison with the figures presented in Sherif et al. (1958), each group's distribution of judgments for the condition it judged first is presented graphically in Fig. 1. As compared with Cond. No-A, composed of the same group of *S*s, Cond. H-A produces a marked assimilation effect, i.e., more frequent use of the heavier categories of judgment for H-A (2nd) than for No-A (1st), for H-A (1st) than for No-A (2nd). To evaluate statistically this assimilation, the percentage of use of each category by each *S* under Cond. H-A was subtracted from his corresponding percentage under Cond. No-A. These percentage differences were analyzed according to a two-factor, Judgment × Order design, different *S*s being used for the two orders but the same *S* having a difference score for each of the six categories. Since the sum of the six difference scores was automatically zero for each *S*, the effects of Order and *S*s were necessarily zero. The Judgment *MS* was highly significant in this analysis ($F = 11.71$, $df = 5/90$, $P < .001$, using as the error term the pooled interaction between *S*s and Judgment—as described in Lindquist, 1953, pp. 267–272). The interaction between Order and Judgment was not significant ($P > .25$, again using the pooled *S*s × Judgment interaction as the error term). This means that the assimilation reported by Sherif et al. (1958) was found to be a reliable effect of the present replication (the distributions of judgments being significantly different for Cond. No-A and H-A). It also means that there is scant evidence of the magnitude of the assimilation being affected by the order in which the two conditions are presented.

However, the distributions of judgments under Cond. No-A, H-A, and Anchor-5, respectively, are each affected by whether or not the given condition has been preceded by other conditions, i.e., whether a given condition was the first, second, or third of those judged by the 10 *S*s following each of the two orders of conditions. Judgment × Order analyses of variance were performed for each of the following three pairs of conditions: No-A (1st) vs. No-A (2nd), H-A (1st) vs. H-A (2nd), and Anchor-5 vs. Anchor-5 (3rd). The percentage-use of each of the categories by each *S* was used for the raw scores. The interaction between Order and Judgment was significant in each of the three analyses ($P < .005$, .025, and .01, respectively). Thus while the present replication confirms the assimilation to the heavy anchor reported by Sherif et al. (1958), it also suggests that their procedure of exposing the same group of *S*s to the various experimental conditions markedly affects the distribution of judgments under any given condition. Examination of the direction of this effect indicates that, in each case, the distribution of judgments tends to resemble the distribution of judgments for the previously presented condition(s). This appears consistent with an adaptation-level description since the respective ALs are pulled toward the value of preceding

ALs. The ALs, defined as one-half the sum of the means of the stimuli judged 3 and 4, respectively, are presented in Table 1.

An analysis of variance based on the percentage scores for Cond. No-A (1st) and Cond. L-A indicated that the difference between these two distributions of judgments was not significant ($F=.90$). This absence of assimilation appears inconsistent with the general trend of the light-anchor effects reported by Sherif et al. (1958). However, they introduced the following experimental variations for this comparison: direct lifting of the weights, fewer and heavier weights, fewer categories, and fewer *S*s. While their own data are not entirely consistent for the different light-anchor conditions they report, there is some indication that they obtained less assimilation to light than to heavy anchors. This suggests that a negative time-order error may be counteracting the postulated central-tendency effect for the light-anchor condition. Negative time-order errors are usually found for lifted weights when the comparison stimuli follow the standard, i.e., the comparison weights tend to be judged heavier than the standard, even though they are physically equal to the standard in weight (Michels & Helson, 1954). Additional evidence, consistent with the assumption that this type of constant error is operating in the present situation, will be presented below under Exp. II.

To test the central-tendency interpretation of the assimilation to the heavy anchor, the PSE for the 141-gm. anchor (i.e., the weight judged equal to the anchor) was determined for both Cond. Anchor-5 and Cond. 7-C by calculating the mean of the upper and lower limens for the anchored category. This could not be done for Cond. H-A since the anchored category (i.e., 6) had no upper limen. The respective limens were first determined by a least squares fit to the *z* scores corresponding to the percentage of times each stimulus was judged above the limen in question (Guilford, 1954, pp. 125–128). The resulting Anchor-5 and 7-Category PSEs were 132 and 135 gm., respectively. Since the anchor was 141 gm. in both cases, the data are consistent with the assumption that there was a negative constant error for this anchor (i.e., in the direction of the mean of the experimental stimuli).

EXPERIMENT II: COMPARATIVE JUDGMENT

While the anchoring instructions used in Exp. I tell *S*s how they are to judge weights which appear equal to the anchoring weight, it is possible that these instructions conflict with the more general instructions for absolute judgment. Insofar as such conflict occurs, the PSEs determined for the Anchor-5 and the 7-Category conditions may not indicate the weight which *S*s would actually perceive as equal to the anchor. In the present experiment, the method of constant stimulus differences is used to obtain a more direct measure of the PSE for the two anchor weights used in Exp. I.

Method

The instructions stated that *S* would be presented weights in pairs and that his task was to judge how heavy the second member of each pair felt in comparison with the first, the standard. Judgments were made in terms of the following categories: much lighter—1, lighter—2, slightly lighter—3, slightly heavier—4, heavier—5, and much heavier—6. These comparative categories yield convenient psychometric functions (Michels & Helson, 1954), and their use in place of the usual restriction to two or three categories increased the similarity between Exp. I and II. In other respects, the former procedure was followed: the same six weights, the same randomization of the comparison presentations in blocks of 48, and the same experimental apparatus. Two new groups of 10 *S*s each were run, one under Cond. H-S (Heavy-Standard, 141 gm.) and the other under Cond. L-S (Light-Standard, 55 gm.).

Results and Discussion

A PSE was obtained for each *S* by tabulating the proportion of his judgments of each stimulus which were either 4, 5, or 6 (i.e., judged "heavier" than the standard), converting these proportions to z scores, and then calculating his PSE by the method of least squares. The mean and *SD* of Cond. H-S PSEs were 124 and 12 gm., respectively, with 9 of the 10 PSEs lower than the 141-gm. standard. The mean of these constant errors is significantly negative ($t = 4.43$, $df = 9$, $P < .005$).

Since 6 of the Cond. L-S *S*s judged the lightest comparison weight to be heavier than the standard (both were 55 gm.) on over half the presentations of this pair and since 4 of these 6 judged every presentation of the remaining comparison stimuli heavier than the standard, it was not appropriate to use the same procedure for determining Cond. L-S PSEs. Linear interpolation (or extrapolation) for the mean judgments of the respective stimuli was used instead. This procedure yielded almost the same PSE for Cond. H-S, 123 gm., and a Cond. L-S PSE of 56 gm. (as shown in Fig. 2 where the two PSEs correspond to mean judgments of 3.5 on the respective psychometric functions). The difference in magnitude between the two constant errors (−18 gm. vs. +1 gm.) is consistent with the assumption, made in the previous discussion, that a negative time-order error was counteracting the effects of the light anchor in Exp. I. Thus the constant errors obtained when the stimuli are explicitly compared with the anchoring weights (the standard for the method of constant stimulus differences) correspond to the shifts in the scales of judgment obtained when the absolute judgments are preceded by these same anchoring weights.

It should be noted that the Cond. H-S PSE is lower than the PSEs obtained for this same weight under the Anchor-5 and 7-Category conditions of Exp. I. This difference may be due to differences between the absolute and comparative response scales. As shown in Table 1, the distributions of judgements for the comparative conditions are much less symmetrical (e.g., very few 5 and 6 judgments for Cond. H-S) than the absolute scales for the same stimuli (e.g., for Cond. Anchor-5). This is a consequence of the difference in the instructions for judging the anchoring weight and the standard, the 141-gm. weight being labeled either 5 or 6 in the anchoring instructions used in Exp. I but only 3.5 in terms of the categories for comparative judgment used in Exp. II. However, the difference in symmetry due to this difference in the instructions for judgment would be reduced by the tendency for *S*s to use their respective categories with fixed proportionate frequencies, symmetrical about the middle of the response scale (i.e., 4, 5, and 6 as often as 1, 2, and 3—as asserted with

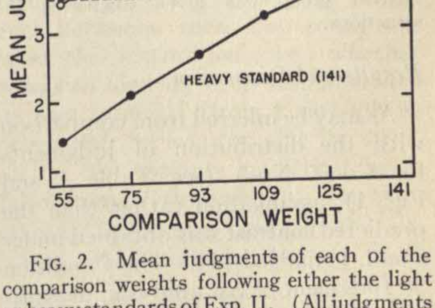


FIG. 2. Mean judgments of each of the comparison weights following either the light or heavy standards of Exp. II. (All judgments were obtained by the method of constant stimulus differences.)

respect to the median-midpoint compromise demonstrated in Parducci, Calfee, Marshall, & Davidson, 1960). The compromise toward a symmetrical scale of judgment would thus produce a lower PSE for Cond. H-S than for either the Anchor-5 or 7-Category conditions. But whatever the value of this interpretation of the differences between the PSEs, it is clear that the PSEs obtained under the comparative instructions are consistent with the adaptation-level interpretation of assimilation.

EXPERIMENT III: UNLABELED ANCHORS

The adaptation-level interpretation of assimilation stresses the anchoring instructions provided by *E*. Without verbal anchoring, the presentation of the 141-gm. weight on alternate trials should produce contrast rather than assimilation (Helson, 1947). Since this extra weight would raise the mean of the stimuli, it would also raise AL and hence lower the judgments.

Method

To test this implication of adaptation-level theory, two additional groups of 10 *S*s each were presented the same series of 576 weights as in Cond. H-A of Exp. I. For Cond. Unjudged-141, *S*s were instructed not to judge the weight presented on every other trial (i.e., the odd-numbered trials); for Cond. Judged-141, *S*s made a judgment on every trial—576 judgments, with no special attention drawn to the alternate trials. Neither group was given anchoring instructions.

Results and Discussion

As may be inferred from comparison with the distribution of judgments for Cond. No-A (see Table 1 and Fig. 1), assimilation rather than the predicted contrast was obtained under Cond. Unjudged-141. The Condition × Judgment × *S*s analysis of variance indicated that the distribution of judgment frequencies for this condition differed significantly from the distribution for Cond. No-A (1st) ($F=3.98$, $df=5/90$, $P<.005$). There is a striking correspondence, however, between the Unjudged-141 and the Anchor-5 distributions. Since tabulation of the 288 odd-trial judgments for Cond. Judged-141 (see Fig. 1) indicates that the 141-gm. weight was judged below 5 almost as often as it was judged 6 (below 5–29%; 5–38%; 6–33%), it appears that the Unjudged-141 *S*s may have anchored their own scales by implicitly labeling each interpolated (i.e., odd-trial) weight as 5 on their 6-point scale.

Why, then, should Cond. Judged-141 yield such a different distribution of judgments (for Judgment × Condition interaction, Judged-141 vs. Unjudged-141, $F = 4.01$, $df = 5/90$, $P < .005$)? Again it seems appropriate to appeal to the postulated tendency to use the categories with symmetrically proportionate frequencies (Parducci et al., 1960). Table 1 shows two different distributions of judgments for Cond. Judged-141. The first is based upon the even trials only (i.e., on 288 judgments); the second is based upon all 576 trials (including the judgments of the odd-trial presentations of the 141-gm. weight). The second is naturally much more asymmetrical (negatively skewed) than either the first or the distribution for Cond. Unjudged-141 since it includes the judgments of 288 extra presentations of the 141-gm. weight (shown in Fig. 1). The hypothesis here is that the difference between the even-trial distributions for Unjudged-141 and Judged-141 reflects the tendency, under the latter condition, to reduce the negative skewing of the total distribution of judgments.

EXPERIMENT IV: IDENTIFICATION

Experiments using absolute judgments have sometimes specified that *S*s were to use their categories of judgment to identify or discriminate the alternative stimuli. Such judgments are *absolute* in the sense that a judgment is made of each stimulus and the implied standards for judgment are not identified during the series of presentations (i.e., the procedure uses the method of single stimuli rather than the method of constant stimulus differences). However, since *S* is told that each category of judgment corresponds to (or identifies) a different stimulus value, each judgment can be scored as either correct or incorrect.

In Exp. I and III, there were six weights and also six categories (except for the 7-Category condition); but this correspondence was never stated explicitly, *S*s being told only that they should judge each weight on a scale from 1 to 6. The difficulty of discrimination was such that *S*s could not learn that there were exactly six weights, and they may not have assumed a one-to-one correspondence between stimuli and categories. Thus the task for judgment in these experiments may resemble closely the task for absolute judgment as established in those experiments for which the number of categories does not correspond to the number of stimuli (e.g., Helson, 1947; Parducci, 1954; Parducci et al., 1960). Such absolute judgments can be neither correct nor incorrect, and the explicit instructions require only that the judgments reflect the psychological relationships among the stimuli. However, the results of recent research have suggested that the use of absolute categories is further restricted by implicit rules of judgment. Thus there is a tendency to use the categories to divide the range of stimuli into proportionate subranges and also to use the categories with fixed relative frequencies (Parducci et al., 1960). Insofar as these implicit rules of judgment do not reflect an attempt simply to identify or discriminate between the alternative stimuli, Exp. I and III might have yielded quite different distributions of judgment if *S*s had been explicitly instructed to identify the stimuli.

Method

To compare the effects of specific instructions to identify the stimuli with the effects of the task for absolute judgment as employed in Exp. I and III, two new groups of 10 *S*s each were told that there were six different weights which they were to identify with the numerals 1 through 6, respectively. Before making any identifications, they were permitted to lift each of the six weights, in random order, without being told which weight was which. The procedure for the Cond. I (Identification) followed the Cond. No-A procedure in all other respects. The procedure for Cond. A-I (Anchored-Identification) was the same as for Cond. H-A except that all six weights served as anchors. Each of the judged weights followed each of the anchors within each successive block of 36 paired presentations. The anchoring weights were correctly identified by *E* on the odd trials, *S*'s task being to identify the weights presented on the even trials.

Results and Discussion

As shown in Table 1 and Fig. 1, the response distribution is more negatively skewed for Cond. I than for Cond. No-A ($F = 6.31, df = 5/90, P < .001$, for the Condition × Judgment interaction). The only difference betweeen these two conditions is in the instructions, i.e., whether *S* was to identify each stimulus with its own numeral (from 1 through 6) or whether he was to judge the stimuli on an absolute numerical scale (also from 1 through 6). The significant difference between the resulting response distributions thus suggests that the absolute scale does not simply reflect *S*'s attempts to identify the stimuli, even when he

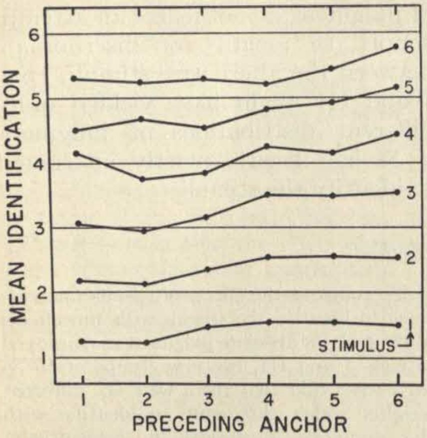


FIG. 3. Mean identification as a function of the preceding anchor for Anchored-Identification condition of Exp. IV.

uses the same number of response categories as there are stimuli.

Paradoxically, *S*s were actually more accurate in identifying the respective stimuli when the correct response for each stimulus was not specified in the instructions (i.e., under Cond. No-A of Exp. I) than under the instructions for identifications (with no interpolated anchoring). The mean percentage of correct judgments was 44% for Cond. No-A (1st), but only 36% for Cond. I ($t = 2.02$, $df = 18$, $P = .05$). The greater accuracy following the Cond. No-A instructions is consistent with the assumption of a tendency to use the categories of judgment with fixed relative frequencies, symmetrical about the middle limen (Parducci et al., 1960). This hypothetical tendency is reflected in the striking symmetry of the Cond. No-A (1st) distribution of judgments as compared with the marked negative skewing found for Cond. I. The skewing of the identifications is presumably due to the negative time-order error demonstrated in Exp. I and II. Thus insofar as the PSEs for the heavier weights are below their actual physical values, *S*s identify them as heavier than they actually are. This is another application of the concept of self-anchoring which was used to explain the assimilation found in Exp. III. The time-order error has less effect following the instructions for Cond. No-A since they place less restriction on *S*'s tendency to distribute his categories with proportionate frequencies.

As might be expected on the basis of the greater amount of information given, accuracy was greater for Cond. A-I than for Cond. No-A (52% vs. 44% correct, $t = 2.37$, $df = 18$, $P < .05$). Separate tabulation was made of the identifications of each stimulus following each anchor for Cond. A-I, and the mean identification of each stimulus is plotted against the value of the preceding anchor in Fig. 3. The slopes of all six of the resulting curves are positive. A trend test, based upon the linear components of these curves, indicated that while the slopes did not differ significantly from each other ($F = 1.57$, $df = 5/45$, $P > .10$, for Stimuli tested against the *S*s × Stimuli interaction), the mean slope

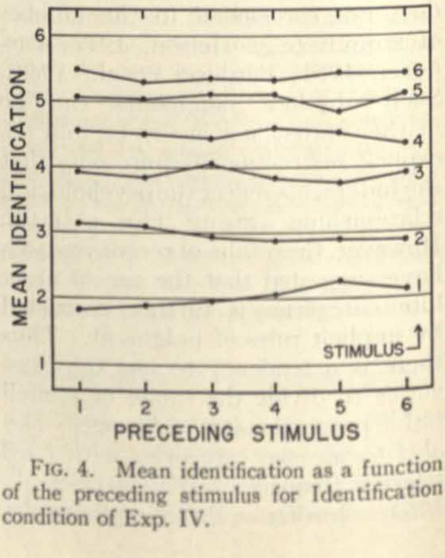


FIG. 4. Mean identification as a function of the preceding stimulus for Identification condition of Exp. IV.

was significantly greater than zero ($F = 24.82$, $df = 1/9$, $P < .001$, for Mean tested against *S*s). It may thus be concluded that the judgment of each stimulus varies *directly* with the value of the preceding anchor. This further case of assimilation is also consistent with the central-tendency interpretation, assuming that the constant errors for the anchor PSEs are more negative for the heavier than for the lighter anchors and that the most recently presented anchor has a greater effect upon judgment than the preceding anchors.

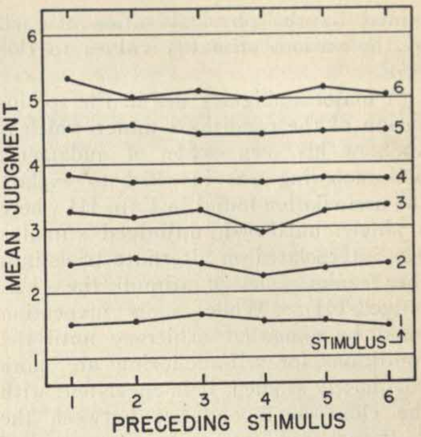


FIG. 5. Mean judgment as a function of the preceding stimulus for No-Anchor condition of Exp. I.

Similar curves have been reported for identification-judgments of loudness made without specific anchoring of the stimuli except during an initial practice session (Garner, 1953). It may be that Garner's situation produced considerable self-anchoring. However, the same tabulation for the present Cond. I and No-A (using, as Garner did, the preceding stimulus value instead of the preceding anchor) revealed no systematic variation in response as a function of the preceding stimulus (see Fig. 4 and 5). Garner proposed two alternative interpretations for his assimilation data. The first was that *S*s confuse the presented stimulus with the preceding one. This would not explain why assimilation was obtained for the present identifications only when they were anchored. Garner's second interpretation was that *S*s tend, when in doubt, to repeat their preceding identifications. The actual mean repetition rate for the unanchored identifications in the present experiment is only 17.8%, not significantly greater than the actual repetition of the stimuli. However, the mean rate of repetition of the preceding anchor category under Cond. A-I is 20.6%, significantly greater than chance ($t = 2.71$, $df = 9$, $P < .025$). Thus the occurrence of assimilation is associated with a tendency toward repetition. A breakdown by category of the Cond. A-I repetitions indicates that the rate of repetition varies systematically with the categories. Insofar as the PSEs for the more extreme anchors are shifted further toward the mean of the stimuli (time-order error being removed by the multiple anchoring), repetition should be greater for the more extreme categories. The mean percentage of the total use of each category represented by repetitions is as follows: 24, 18, 19, 19, 20, and 29% for 1 through 6, respectively. Although statistical evaluation of these differences is precluded by the variation in the denominators (the actual frequency of use of the respective categories), the curvilinear form of the repetition function is consistent with the hypothesized central-tendency effect.

DISCUSSION

These experiments demonstrate that assimilation effects are found for a variety of simple judgmental tasks. However, in each of the present cases, the assimilation is consistent with the contrast principles described by the theory of adaptation level. Whenever a category of judgment is anchored to a specific stimulus, either by *E* or by *S*, the anchored category is applied to the PSE for the anchoring stimulus in accordance with the adaptation-level analysis for comparative judgment. The remaining judgments are largely determined by the physical ratios of each of the various stimulus values to this PSE.

A major ambiguity lies in the specification of the conditions under which *S* anchors his own scale of judgment. Self-anchoring was invoked to explain the assimilation found in Exp. III where a single, unlabeled, unjudged stimulus was interpolated on alternate trials into the regular series of stimuli (i.e., Unjudged-141). While such invocation must be somewhat arbitrary until the conditions for self-anchoring are more adequately studied, it is consistent with the close correspondence between the "self-anchored" scale of judgment and the scale anchored with the category which *S*s would otherwise apply to the anchoring stimulus (i.e., Anchor-5). The adaptation-level approach, even with this ambiguity, would appear to provide a more powerful analysis than the common use of assimilation and contrast as two, basically opposed principles of judgment, capable between them of describing any possible shift in judgment.

While the results of these four experiments provide a comparison between various kinds of absolute and comparative judgments, the analysis stresses the anchoring features common to all of the procedures studied. The net effect is to emphasize the role of comparative judgment, comparison between the judged stimulus and the values of previously presented standards or anchors. This suggests that instead of the common distinction between absolute and comparative judgments, it would be more appropriate to classify the various procedures for judgment with respect to the manner and degree to which the rules for judgment have been specified.

Unanchored absolute judgments, elicited by the customary instructions to use verbal categories (e.g., light, medium, heavy), would appear to be the least specified, the least anchored by *E*. The numerical scale used for Cond. No-A of Exp. I approximates this lowest degree of specification, and the distribution of judgments under this condition is similar to the distribution obtained using verbal categories, i.e., symmetrical for a rectangular distribution of stimuli, with more frequent use of the middle categories (see Parducci et al., 1960, p. 71). But even though the instructions for such scales involve little explicit anchoring, the categories of judgment are to a considerable extent anchored by implicit rules for judgment. Thus each of the No-Anchor *S*s used all six categories; and as indicated in Fig. 5, the order of the categories corresponds to the order of the stimuli. As has been shown in previous experiments, the addition of an unlabeled weight which was heavier than 141 gm. would have anchored Category 6 to the new value, producing a contrasting shift in the judgments of the series stimuli (e.g., Postman & Miller, 1945). Much of the research on absolute judgment has been directed toward the discovery of such implicit rules of judgment.

When individual stimulus values are not singled out for special attention, *S*s have greater difficulty in applying either the implicit rules of judgment or the anchoring rules provided by *E*. Thus identifications were actually less accurate following the instructions stating the one-to-one correspondence between categories and stimuli than following the No-Anchor instructions. The Identification *S*s had no constant physical landmarks to which they could anchor their categories, and it appeared that the instructions to identify conflicted with the implicit rule to use the categories with fixed relative frequencies.

The introduction by *E* of a constant physical landmark, such as the unlabeled anchor interpolated into the series for Cond. Unjudged-141 of Exp. III, provides a further degree of specification. Insofar as this stimulus is recognized as unchanging, the category which *S* applies to it is anchored at a fixed value of the stimulus. Explicit anchoring of this stimulus by *E*, as in any of the anchoring conditions of Exp. I, would completely determine the location of at least one point on the scale of judgment if it were not for the changes in PSE for the anchored value. Similar constant errors (systematic shifts in PSE) were found following both the absolute and comparative instructions, though the shift in PSE for the 141-gm. weight was actually less using the absolute procedure. Again, the implicit rule to distribute the categories with proportionate frequencies may have produced the greater accuracy of the absolute scale.

The Anchored-Identification condition involved the most complete specification of the various categories of judgment. Although the generally negative time-order error was effectively eliminated under this condition, the assimilation toward the values of the preceding anchors (shown in Fig. 3) appeared consistent with the adaptation-level analysis of the constant errors in the PSEs for the anchored values. Thus the same sort of perceptual contrast was obtained when all six categories were repeatedly anchored to their respective stimuli as when there was less specification of the rules for judgment. The conclusion from this analysis is that while the nature and degree of specification determine what shifts in judgment will occur, the judgments are determined by the physical relationships between the judged stimuli and the PSEs for the anchoring or standard stimuli.

SUMMARY

A successful replication was described for a recent demonstration of assimilation, rather than the usual contrast, in the anchoring of lifted weights. This assimilation was interpreted as a special case of the central-tendency effect, with the anchor serving as the standard for comparative judgments. Using both absolute and comparative judgments, it was shown that the PSE for the anchor shifted toward the center of the stimulus series in accordance with the contrast effects described by the theory of adaptation level. While it was necessary to invoke self-anchoring to explain assimilation to an unlabeled stimulus, this permitted reduction of all cases of assimilation to simple perceptual contrast. The analysis suggested that various procedures for absolute and comparative judgment are distinguished primarily on the basis of the nature and degree of specification of the rules for judgment.

REFERENCES

GARNER, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46**, 373–380.

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

HELSON, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of science.* Vol. 1. New York: McGraw-Hill, 1959.

HOLLINGWORTH, H. L. The central tendency of judgment. *J. Phil. Psychol., sci. Meth.*, 1910, **7**, 461–468.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

LONG, L. A study of the effect of preceding stimuli upon the judgment of auditory intensities. *Arch. Psychol., N. Y.*, 1937, **30**, No. 209.

MICHELS, W. C., & HELSON, H. A quantitative theory of time-order effects. *Amer. J. Psychol.*, 1954, **67**, 327–334.

PARDUCCI, A. Learning variables in the judgment of single stimuli. *J. exp. Psychol.*, 1954, **48**, 24–30.

PARDUCCI, A., CALFEE, R. C., MARSHALL, L. M., & DAVIDSON, L. P. Context effects in judgment: Adaptation level as a function of the mean, midpoint, and median of the stimuli. *J. exp. Psychol.*, 1960, **60**, 65–77.

POSTMAN, L. J., & MILLER, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58**, 43–53.

ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol., N. Y.*, 1941, **37**, No. 261.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

(Received February 3, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 438–443

CUE NEUTRALIZATION: THE EFFECTS OF RANDOM REINFORCEMENTS UPON DISCRIMINATION LEARNING[1]

MARVIN LEVINE

Indiana University

The present experiments are concerned with the effects of random reinforcements (RRs) upon subsequent discrimination learning. Human *S*s are presented with a sequence of two colored stimuli (e.g., blue or tan cards), are instructed to make one of two responses to each stimulus, and are told either "right" or "wrong" following each response. During a first procedural phase, for a given number of trials, *E* applies the reinforcing events on a prearranged random schedule, i.e., the words "right" and "wrong" are uncorrelated with any stimulus-response pairs. During a second phase, which appears without special announcement or break, the traditional discrimination procedure is employed in which the word "right" is presented only when the correct response is made to the appropriate color. Otherwise the word "wrong" is presented. One may then inquire into the effect of the preliminary RRs upon the subsequent discrimination learning by evaluating the difficulty encountered in learning the discrimination problem.

The experimental results will have implications for the process of cue neutralization, and for theories which describe this process. Restle (1955; Bourne & Restle, 1959) and Bush and Mosteller (1951) have, in describing discrimination learning, included the idea that cues which are uncorrelated with the reinforcement lose control over the responses. Specific predictions may be made for the results of the present experiments. These predictions and their relationship to the outcome of the present experiment will be described following the presentation of the results.

EXPERIMENT I

Experiment I served to obtain broad coverage on the variable of number of RRs. Either 0, 10, 30, or 60 RRs were presented before the discrimination problem began.

A second variable resulted from consideration of a problem first noted by Spence (1945). It is possible that *S* could literally ignore the stimuli by "closing his eyes" to them. As a result of the RR *S* might decide that the colors had nothing to do with the problem and might make his responses without adequate receptor stimulation, i.e., without directing his gaze to the colors. The decrement in learning which would subsequently appear would in a sense be an artifact, since the reinforcements, when they came regularly, could not strengthen any S-R relation. In order to evaluate whether such an artifact was occurring half of the *S*s were forced to observe the stimuli. They were instructed to describe each stimulus before making their response. The other half of the *S*s did not receive this special instruction. If RRs retarded performance only in the latter

---

[1] This research was supported in part by Research Grant M-3906A from the United States Public Health Service. Thanks are due Samuel H. Revusky and Charles Land for serving as *E*s for half the *S*s.

group, it would suggest, as one possibility, that these *S*s were not always viewing the colors.

Method

Subjects.—The *S*s were 16 Indiana University students in an introductory psychology course taught by the author.

Apparatus.—Four decks of cards were employed, one for each of four problems. Within each deck one of two forms (e.g., a diamond or cross) appeared on each card, and each form was of one of two colors (e.g., purple or tan). The four resulting stimuli were distributed randomly throughout a deck of 120 cards.

The four decks were composed of the following stimuli: a purple or tan diamond or cross; a red or green square or semicircle; a yellow or blue circle or X shape; an orange or black triangle or star.

Design.—The observing-instruction variable and 2 different *E*s were organized factorially to produce four different conditions. Four *S*s were assigned to each of the four conditions. Each *S* received four discrimination problems preceded by either 0, 10, 30, or 60 RR trials. The 4 *S*s within any one condition received the four RR amounts in four different sequences counterbalanced to produce a latin square. A different latin square was employed for each condition.

Procedure.—Each *S* was run alone by 1 of the 2 *E*s. The *E* presented *S* with one of the four decks face down, and showed him four different sample cards which were the types making up the whole deck. The *S* was instructed to turn the cards in the deck one at a time and to "sort" the cards into two groups by making one of two responses to each card. The two responses were given as VEC and NONVEC. The *S* was further told that after each response *E* would tell him whether he was right or wrong and that his task was to solve the problem in the sense of getting as many right as possible. If *S* was in the observing-instruction condition he was also told to state the color and form on the card before making his response.

The experiment then proceeded in two main phases: Phase I, the RR part of the experiment, and Phase II, the discrimination proper. Following each response in Phase I, *E* reinforced according to a previously constructed schedule of "rights" and "wrongs." The two events were in a sequence determined randomly with two restrictions: there were never more than three "rights" (or "wrongs") in succession, and half the outcomes were always "rights."

When Phase I was completed *E*, without comment, or obvious change in behavior, shifted his basis for reinforcement to *S*'s responses to the cards. The particular stimulus-response pairs receiving the word "right" had been designated in advance. For example, the positively reinforced pairs might be purple = VEC, tan = NONVEC. Phase II, therefore, consisted in a two-response discrimination problem with one relevant dimension (color) and one irrelevant dimension (form) and with 100% consistency in the stimulus-response-reinforcement combinations. The problem continued in this way either until *S* reached a criterion of 15 consecutive correct responses or, if criterion performance did not appear, until 50 responses had occurred. Following the attainment of the criterion *E*, again without providing any discriminable cue, presented RRs for the next 10 responses. The purpose of this last detail was to minimize transfer of a set for simple solutions from one problem to the next.

The deck was then collected and replaced by the deck for the second problem. The *S* was instructed that the second problem followed the same procedure as the first but was otherwise a brand new problem having no relation to the preceding. The four sample cards for the new deck were shown and two new responses (ZEB and NONZEB) given. The *S* was again told to get as many right as possible, and, if he was in the observing-instruction condition, was asked to describe the color and form on the card before making his response.

Following Phase II of this second deck, the third and fourth decks followed in sequence with instructions similar to those preceding the second deck.

Results

Trials to criterion and number of errors were taken as measures of performance during the discrimination problem. Both measures yielded similar results and the same conclusions. Only the trials to criterion measure, therefore, will be presented here.

The effects of different numbers of RR trials and the influence of the observing-instruction variable upon these effects are shown in Fig. 1.

Two general results may be seen concerning the RR variable. The 0 condition produces faster learning than the other three conditions and the function appears to be flat between 10 and 60 RRs. This graph suggests that the changes produced by RRs occur during the first 10 RR trials and that beyond this number further RRs do not impede learning. Figure 1 also shows that the observing-instruction variable has little effect, and that what effect it does have is in the direction opposite from that expected by the consideration that the noninstructed *S*s might not look at the stimuli. When *S*s are forced to describe the stimulus before each response, thereby demonstrating that the stimulus has been observed, the performance is slightly worse.

The effect of the observing instruction was evaluated statistically. Total trials to criterion following all four RR amounts were obtained for each *S*, these totals were treated as the cell scores in the 2 × 2 factorial design produced by the *E* and instruction variables, and an analysis of variance was performed on these data. Neither the *E*, the observing-instruction variable, nor the interaction was significant.

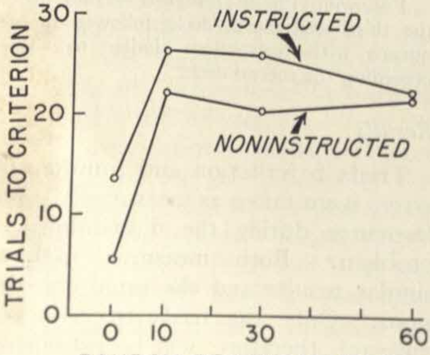


FIG. 1. Mean number of trials to reach criterion following different numbers of random reinforcements, for the group instructed to describe the stimuli and the noninstructed group.

EXPERIMENT II

Experiment I indicated that the value of 10 RRs yielded results which were already at asymptote, that if one wished to show a gradual increase in problem difficulty with an increase in number of RRs, one had to present small numbers of RRs. Experiment I also utilized a latin square design. A problem exists with this design in that the resulting function may describe not only the effects of the variable but may also be influenced by the particular sequences in which the levels of the variable were presented. Accordingly, in Exp. II independent groups were compared in which *S*s received a single problem preceded by either 0, 4, 8, or 12 RRs.

The effects of irrelevant dimensions were also investigated. In Exp. I there was always one relevant (color) and one irrelevant dimension (form) during the discrimination problem. In Exp. II there was again one relevant dimension (color), but different groups received problems with either 0, 1, or 2 irrelevant dimensions.

Method

Subjects.—The *S*s were 144 Indiana University students in an introductory psychology course taught by the author. The *S*s were largely freshmen and were not from the same class as *S*s of Exp. I.

Apparatus.—More than 1,000 3 × 3 in. cards were printed, each containing a single figure. The figure was either a diamond or a cross, which could be either blue or tan and large (1 sq. in.) or small (¼ sq. in.). These cards were divided up into nine decks of 120 cards. All the decks varied with respect to color in exactly the same way: In each set of 20 cards blue and tan each appeared 10 times; neither color appeared on more than three successive cards. The nine decks differed with respect to the number of noncolor dimensions which were present, as follows: (*a*) 0-decks: There were four different decks which did not vary in any dimension other than color. In one deck, for example, all the stimuli were large diamonds. The other three decks consisted of small diamonds, large crosses, and small crosses, respectively. (*b*) 1-decks: There were four different decks which varied in one dimension independently of the variation in color. In one deck, for example, all the stimuli were large diamonds or crosses. The other three decks consisted of small diamonds or crosses, large or small diamonds, and large or small crosses. (*c*) 2-deck: There was one deck which varied in two dimensions independently of the variation in color. The stimuli could be large or small crosses or diamonds.

All the irrelevant dimensions had the same sequential characteristics in all decks: Each figure (or size) appeared 10 times in 20 cards; no more than 3 cards in succession had the same figure (or size).

Design.—The variables of number of RR trials (0, 4, 8, or 12) and number of irrelevant dimensions (0, 1, or 2) were arranged factorially to produce 12 different conditions. Twelve different *S*s were used in each condition. Within each of the 0 and 1 irrelevant-dimension conditions 3 of the *S*s each received one of the four available decks.

Procedure.—The *E* presented *S* with the appropriate deck face down and showed him the different cards making up the deck (two cards in the 0-decks, four cards in the 1-decks, and eight cards in the 2-deck). The *S* received the same instructions as *S*s in Exp. I. No special observing instructions were given to any *S*, however.

Following each response in Phase I *E* reinforced according to a previously constructed schedule of "rights" and "wrongs." The two events were in a sequence randomly determined with the same restrictions as in Exp. I. An additional restriction was that the last response in Phase I was always called "wrong."

The shift to the Phase II procedure was without detectable change in *E*'s behavior. The relevant dimension was always color. Phase II ended either when *S* reached a criterion of 15 consecutive correct responses or when 100 responses had been made, whichever was less.

Results

The overall mean trials to criterion are shown in Fig. 2. The preliminary RR trials produce the same two effects that appeared in Fig 1: The 0 condition produces faster learning than the other three conditions, and these latter conditions do not appear to differ among themselves. These results are reflected in the statistical tests. Analysis of variance of all 12 cells shows that the number of RRs is significant. Analysis of the 9 cells of the nonzero conditions shows that the RR variable is not significant.

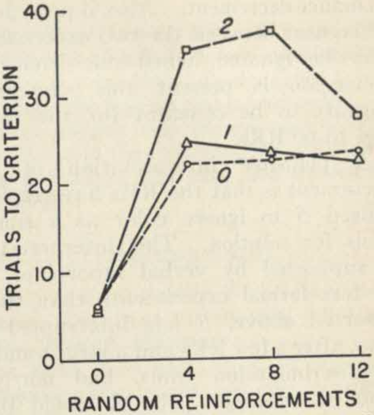


FIG. 2. Mean number of trials required to reach criterion following different numbers of random reinforcements. (The parameter is the number of irrelevant dimensions present during discrimination learning.)

Figure 2 also suggests that the number of irrelevant dimensions affects 0 RRs differently from the other amounts of RR. The 0, 1, and 2 irrelevant dimensions do not appear to produce differences under 0 RRs, whereas 2 irrelevant dimensions appear to produce greater difficulty for the *S*s receiving 4, 8, or 12 RRs. It should be added, however, that these results are not corroborated by the statistical tests: The interaction between RR and irrelevant dimension is not significant.

DISCUSSION

Figures 1 and 2 show that 4 or more RRs produce a considerable mean performance decrement. Also, if procedural differences between the two experiments may be ignored, when one irrelevant dimension is present this decrement appears to be constant for the range of 4 to 60 RRs.

A plausible interpretation of the decrement is that the RRs have rapidly caused *S* to ignore color as a simple basis for solution. This interpretation is supported by verbal reports of *S*s. In less formal experiments than those reported above, *E* has interrupted *S*s who, after a few RRs and a large number of discrimination trials, had not yet reached criterion. The *E* would then ask *S* what he thought was involved. The *S* invariably indicated that he was attending to a complex sequence (e.g., "say VEG to blue twice then say NONVEG; do just the opposite for tan," or, "it is going VEG, VEG, NONVEG, VEG, NONVEG, but it sometimes changes"). When *E* pointed out that all blue cards were NONVEG, all tan cards were VEG, *S* frequently remarked that he had tried that at the beginning but that it hadn't worked.

The *S*, then, would appear to be ignoring color as a direct basis for solution. The process by which an *S* comes to ignore cues has been treated, in stimulus sampling theory, as a process of cue neutralization. Serious problems occur, however, when this process, as it has been embodied in recent theorizing, is applied to the present results.

Restle (1955) characterized a neutralized cue as "one which the subject does not consider in deciding upon his choice response" (p. 12). Cues, according to his model, gradually become neutralized when they are irrelevant, i.e., when the reinforcements are uncorrelated with the presence or absence of particular responses to those cues. He defined the process of cue neutralization (or, in his earlier terminology, cue adaptation) with this sentence: "If $a(K, n)$ is the probability that irrelevant cue K has been adapted at the beginning of the nth trial, then:

$$a(K, n+1) = a(K, n) + \theta[1 - a(K, n)]$$

is the probability that it will be adapted by the beginning of the next trial" (Restle, 1955, p. 12).

The equation shows that neutralization depends upon the parameter, θ, which is defined as the proportion of relevant cues (cues which, according to Restle, can be used by the subject to predict where or how reward is to be obtained). Its role in the equation is such that cues do not become neutralized when $\theta = 0$, a condition which holds when there are no relevant cues. During the RR phase, in the experiments reported here, color is an irrelevant cue (reinforcements are random with respect to responses made to the colors) and $\theta = 0$ (there are no relevant cues because of the random character of the reinforcement sequence). Because of the latter feature the model predicts that color will not become neutralized, that *S*, after a large number of RRs, will be no more likely to ignore color than after 0 RRs. Figures 1 and 2, however, clearly reveal a neutralization effect.

The same idea is also suggested by Bush and Mosteller (1951). The operator, *D*, which may be interpreted as the mathematical analogue of stimulus neutralization, does not exist when the discriminanda are indiscriminable, i.e., during a random reinforcement series. This operator is assumed only when relevant cues are present. The present results do not confirm this formulation. They suggest, rather, that the models will have to be revised to permit neutralization to be independent of cue relevance.

Performance during Phase II poses a second problem for these theories, for they assume that the process of cue neutralization is irreversible. Once cues are neutralized they are permanently ignored, i.e., they are no longer available for conditioning. Experiment II implies, however, that the color cues are not permanently neutralized. Of the 108 *S*s receiving four or more RRs, only 1 failed to solve the discrimination problem within 100 trials.

In short, the present results suggest that cues are neutralized independently of the existence of relevant cues, and that the sampling probability of a neutralized cue, while it may become reduced, does not necessarily become zero.

SUMMARY

The effects of random reinforcements (RRs) upon discrimination learning were investigated in two experiments. Two colors were presented in a random sequence and *S* made one of two responses each time a colored stimulus appeared. The *E* said "right" or "wrong" after each response, first in a prearranged sequence (the RR phase), then contingent upon the occurrence of particular stimulus-response pairs (the discrimination phase). In Exp. I the effects of 0, 10, 30, and 60 RRs upon discrimination learning were compared. Experiment II compared the effects of 0, 4, 8, and 12 RRs. Random reinforcements always produced a decrement in performance on the discrimination problem. The decrement appeared to be constant from 4 to 60 RRs. In contrast to the cue-neutralization process described by mathematical models of discrimination learning, the results suggest that cue neutralization does not depend upon the existence of relevant cues, and that the sampling probability of a neutralized cue does not necessarily become zero.

REFERENCES

BOURNE, L. E., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

SPENCE, K. W. An experimental test of the continuity and noncontinuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253–266.

(Received February 6, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 444–451

REVERSAL AND NONREVERSAL SHIFTS IN DISCRIMINATION LEARNING IN RETARDATES[1]

BETTY HOUSE AND DAVID ZEAMAN

Mansfield State Training School *University of Connecticut*

The second of two discrimination problems may be learned more quickly if the same cues are relevant in both than if different relevant dimensions are presented. This effect has been shown when the possibility of positive transfer of instrumental responses has been ruled out (Lawrence, 1949, 1950) and even, in some cases, when negative transfer of instrumental responses is expected, as in reversal of discriminations (Buss, 1953, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955; Kendler & Kendler, 1959).

The "observing response" model of Wyckoff (1952, 1954) might account for these findings. According to the theory, probability of observing relevant cues increases as a discrimination is learned while probability of observing irrelevant cues decreases. Consequently, relevant cues in a second problem are more likely to be observed if they are the same than if they are different from those in the first problem. The relative frequency of effective over ineffective training trials is thereby increased, and learning is faster when the same cues are relevant in both problems.

The model is thus consistent with the finding that sometimes reversal is learned faster than a new problem with stimuli from a different relevant dimension. A further test is suggested if one assumes, as have Kurtz (1955) and House and Zeaman (1960), that "observing responses" have a broad topography, i.e., may be specified as observation of any value of an entire stimulus dimension such as color or size. It follows that facilitation would also be obtained if the second problem has different relevant cues, but from the *same* relevant dimension.

Some predictions about the expected shapes of learning curves during these transfer tests can also be derived from the theory. Since an observing response must be made to the relevant cue before any learning can occur, an initially low probability will likely produce a number of trials at chance performance preceding the first effective training trial. The result would be the commonly observed ogival discrimination learning curve. Experimental operations designed to decrease or increase initial probability of observing the relevant dimension should have the specific effect of lengthening or shortening, respectively, the initial flat portion of the learning curve.

In the present study, these predictions are tested by training retarded *S*s with a single relevant dimension, then testing under three different shift conditions: (*a*) reversal, (*b*) new problem with cues from the same relevant dimension (intradimensional shift), and (*c*) new problem with cues from a different relevant dimension (extradimensional shift). In addition, two training conditions are employed in which irrelevant cues are either variable or constant. Both training methods have been used by previous investigators (e.g., Kendler & Kendler, 1959; Kendler, Kendler, & Wells, 1960), but no comparison has been made to determine if there are differential effects on performance during the shift tests.

Under the Variable Irrelevant (Variable Irrel.) training condition, two objects differing in color *and* form appear on each trial, e.g. red square vs. green circle. Form might be relevant with the square as positive cue. The square and circle exchange colors randomly from trial to trial. Following such training a group of *S*s is next split into matched groups to enter the three shift conditions.

---

[1] This investigation was supported by Research Grant M-1099 from the National Institute of Mental Health, Public Health Service. The authors wish to express appreciation to Neil A. Dayton, Superintendent, and to John Cassell and Louis F. Boly, Director of Training and Principal of the Longley School, for their aid and cooperation.

Reversal (Rev.) training requires no change in the stimuli. The circle would now be positive instead of the square. Intradimensional Shift (Intra-D Shift) keeps form relevant but shifts to new form cues such as triangle and cross. The old irrelevant colors, red and green, are randomly associated with triangle and cross. Extradimensional Shift (Extra-D Shift) changes the relevant dimension. In the example above, the originally irrelevant red and green cues now become relevant and these randomly exchange two shapes not used in training, such as triangle and cross.

These training and shift conditions provided the experimental design for groups of retardates of low mental age (MA 4–6 yr.). Equivocal results led to a replication with higher MA groups and an added training condition with constant irrelevant stimuli. For the Constant Irrelevant (Constant Irrel.) training condition, the two objects differ in only a single dimension, either color *or* form. An example corresponding to Variable Irrel. above would be a red square vs. red triangle. A reversal was not run following Constant Irrel. training. Intra-D and Extra-D Shift conditions are identical to those after Variable Irrel. training.

METHOD

Subjects

All *S*s were institutionalized, mentally retarded boys and girls attending classes at the Longley School of the Mansfield State Training School of Connecticut.[2] Random selections were made within given MA ranges without regard to diagnosis or previous experience. Three training groups may be distinguished as follows:

Variable Irrelevant: MA 4–6.—Thirty-seven *S*s, 23 boys and 14 girls, with mean MA of 59 mo., mean IQ 41 (range 26 to 62), and mean CA 151 mo. (range 79 to 216), comprised the group.

Variable Irrelevant: MA 6–8.—In this group were 30 *S*s, 17 boys and 13 girls, with mean MA of 83 mo., mean IQ 56 (range 38 to 69), and mean CA 151 mo. (range 121 to 196).

Constant Irrelevant: MA 6–8.—In this group were 20 *S*s, 10 boys and 10 girls, with mean MA 85 mo., mean IQ 55 (range 39 to 73), and mean CA 161 mo. (range 104 to 233).

Apparatus

A modification of the Wisconsin General Test Apparatus was used. Components were (*a*) a sliding tray containing two 3-in. food cups centered 12 in. apart and (*b*) a one-way screen separating *E* from *S*. There were 25 different stimulus objects consisting of all combinations of five colors: red, green, yellow, black, and white; and five forms: triangle, square, circle, T, and cross. Forms were cut from ¼-in. Masonite and mounted vertically on 4 × 4 in. gray Masonite bases. Maximum height and width of each stimulus was 2 in.

General Procedure

To bait the food cup, *E* pulled the sliding tray behind the one-way screen, then pushed the tray directly in front of *S* to begin a trial. The *S*'s response was lifting the stimulus object from the concealed cup. Candy rewards were M & Ms, candy corn, and the like. The *E* said "Good" if the first response was correct, and "No" if it was not. Immediate correction of wrong responses was allowed. Position of the correct stimulus, left or right, was varied according to a Gellermann series.

For both training and shift tests, 25 trials per day were given. Criterion of learning was 20/25 correct during a single daily session. For the original problem, training was continued for 125 trials of overlearning beyond criterion. Failure criterion was 250 trials without reaching criterion.

Training

Within each training group, *S*s were assigned randomly, half to a color-relevant problem and half to a form-relevant problem.

---

[2] Intelligence test scores were from the Stanford-Binet, 1916 revision, provided by the Psychology Department of the Mansfield State Training School, under the direction of Charles Fonda.

Variable Irrelevant.—From the five colors and five forms available, each *S* was randomly assigned a pair of relevant cues from one dimension, and a pair of irrelevant cues from the other. Thus, a particular *S* might have red and green as relevant cues, square and triangle as irrelevant. To make the irrelevant cues uncorrelated with reward for this *S*, on half the trials the positive cue is associated with one of the two irrelevant cues, and on the remaining trials with the other irrelevant cue. A red triangle vs. green square is alternated over trials with red square vs. green triangle according to a Gellermann series.

Constant Irrelevant.—From either the five colors or the five forms, each *S* was randomly assigned a pair of relevant cues. From the remaining dimension a single cue was randomly chosen. A particular *S* might then have a yellow circle vs. a yellow square for a form relevant problem.

Special Training

For *S*s in the two MA 6–8 training groups, special training was introduced after 250 trials without learning. This procedure was necessary because of a high failure rate. The alternative of training more *S*s and using only learners for the shift tests might produce an artifact. Suppose that *S*s are of two types, those who attend mostly to form and those who attend to color. With a high proportion of failures, there is a possibility that most learners are simply those assigned by chance to their dominant dimensions. Then difficulty upon being shifted to another dimension could be attributed to the introduction of a weaker cue rather than to the effects of training per se. Of the 30 *S*s in the Variable Irrel. group, 12 required special training, as did 10 *S*s of the 20 in the Constant Irrel. group. Since all eventually learned, the possibility of this artifactual explanation is ruled out.

Procedure A.—After 250 failure trials an object differing multidimensionally from the positive cue was selected from a large collection of "junk" objects and sustituted for the negative cue. After criterion, 20/25 correct during a daily session on this simplified problem, regular training procedures were resumed with the original stimuli. Failure criterion was 100 trials for each stage without getting 20/25 correct. Of 22 *S*s receiving this training, 9 continued to fail and were given the following training.

Procedure B.—Without mentioning the color or form of the cues, *E* showed the positive stimulus objects to *S* saying, "These blocks are always right. The candy is always under these blocks." The two incorrect blocks were then presented by *E* with the words, "These blocks are always wrong. The candy is never under these blocks." The appropriate modification was made for the Constant Irrel. group since only one positive and one negative stimulus were used. After these verbal instructions, regular training was begun with no further help from *E*. All *S*s then met criterion of 20/25 correct and received 125 trials of overlearning.

Shift Conditions

The general procedure for the shift tests was the same as for training. The color-relevant and form-relevant training groups were evenly divided among the three shift conditions. The three shift groups were approximately equated for speed of original learning. In particular, *S*s receiving special training were equally represented under each shift condition.

Reversal.—Only *S*s from the Variable Irrel. training condition were given the Reversal test. The same stimuli were used for both training and test. Only the reward values were reversed.

Intradimensional Shift.—The *S*s having Variable Irrel. training were assigned randomly a new pair of relevant cues combined with the same pair of irrelevant cues. The *S*s having Constant Irrel. training were assigned two new relevant cues to be combined with two new irrelevant cues replacing the single cue present during training. Both training groups had the irrelevant cue varied during the shift test.

Extradimensional Shift.—For *S*s having Variable Irrel. training, the pair of cues which were irrelevant during training were made relevant. Two new cues were randomly assigned from the previously relevant dimension to become irrelevant during the shift test. For *S*s having Constant Irrel. training, two new cues from the previously irrelevant dimension were assigned as relevant cues, to be combined with two new cues from the previously relevant dimension.

RESULTS

Variable Irrelevant: MA 4–6.—Only 16 of the 37 *S*s in this group learned the original problem within 250 trials. In spite of small *N*, differences among the shift conditions were significant.

Geometric mean errors to criterion were 6.3, 6.6, and 72.5 for Rev., Intra-D Shift, and Extra-D Shift, respectively, ($F=5.11$; $P<.05$). The difference between Rev. and Intra-D Shift performance was not significant ($t < 1$), but the Extra-D Shift group was inferior to both of the other groups ($t = 2.6, 2.7$; $P < .05$). These analyses were carried out using log transforms of the error data.

Shift effects for the MA 4–6 group, though significant, were not conclusive because color and form were not equally represented among the shift conditions. Only 4 *S*s (out of 19) learned an original color problem, while 12 (out of 18) learned a form problem. As a result, both the Rev. and Intra-D Shift groups were mostly *S*s remaining on a form problem for the shift, while those in the Extra-D Shift group were shifting mostly from form to the more difficult color problem. The ambiguity of these results led to a replication of the experiment with brighter *S*s and the special training methods required to ensure an unbiased sample on the shift tests.

Variable Irrelevant: *MA 6–8*.—The ordering of conditions for the MA 4–6 group was reproduced with the MA 6–8 group, with a larger N and an equal number of *S*s trained with color and form relevant under each shift condition. Geometric mean errors were 6.2, 4.4, and 23.2 for Rev., Intra-D Shift, and Extra-D Shift groups, respectively. Table 1 shows these results broken down by cues. A two-way analysis of variance (Shift × Cue relevant during shift) of log errors to criterion showed the three shift groups differed significantly ($F = 4.79$; $P < .05$). Neither cue during shift nor interaction of cue with shift produced significant effects. Again, the Extra-D Shift group performed more poorly than the Rev. group ($t = 2.38$; $P < .025$) and the Intra-D Shift group ($t=3.00$; $P<.01$), but the Rev. and Intra-D Shift groups were not different ($t < 1$). The three groups did not differ significantly with respect to speed of original learning ($F = 1.02$; $P > .20$). Geometric mean errors to criterion or (for *S*s requiring special training) during 250 trials of failure were 30, 38, and 16, respectively, for the Rev., Intra-D Shift, and Extra-D Shift groups. The 12 *S*s requiring special training were evenly divided among the three shift conditions, and their performance scores showed the same effects as the total group.

TABLE 1

GEOMETRIC MEAN ERRORS TO CRITERION FOR 10 GROUPS OF *S*S DURING 10 DIFFERENT TRANSFER TESTS

Irrelevant Cue during Training	Conditions during Transfer Tests					
	Reversal		Intra-D Shift		Extra-D Shift	
	Form (F)	Color (C)	F→F	C→C	F→C	C→F
Variable	9.6	4.0	3.1	6.3	20.4	26.3
Constant	—	—	1.8	11.5	44.7	2.0

Constant Irrelevant, *MA 6–8*.—The difference between the Extra-D Shift and Intra-D Shift groups after constant irrelevant training was not found significant ($F=1.17$; $P>.25$). The geometric means of the two shift conditions were 4.6 and 9.3 errors to criterion for Intra-D Shift and Extra-D Shift, respectively. Table 1 shows the results for color and form separately. A two-way analysis of variance (Shift × Cue relevant during shift) of log errors to criterion showed that color was more difficult than form ($F = 14.12$; $P < .005$), but interaction of Cue × Shift was not

TABLE 2

Analysis of Variance of Log Errors to Criterion during Shift for Intra-D Shift and Extra-D Shift after Variable Irrel. and Constant Irrel. Training

Source	*df*	*F*
Shift (S)	1	7.53**
Training (T)	1	1.02
Cues (C)	1	9.91**
S × T	1	1.17
S × C	1	<1
T × C	1	6.82*
S × T × C	1	1.69
Error (*MS*)	32	(.3530)

* $P < .05$.
** $P < .01$.

significant ($F < 1$). Differences between the shift groups during original learning were negligible. Geometric mean errors to criterion or 250 failure trials were 61 and 53 for Intra-D Shift and Extra-D Shift, respectively. The 10 *S*s who required special training were evenly divided among the shift conditions and their results were in the same direction as for the total group.

Constant vs. Variable Irrelevant training.—As a further analysis of the effects of training a 2 × 2 × 2 factorial analysis of variance was performed on the scores of the Intra-D Shift and Extra-D Shift groups for both training conditions. The factors were type of shift (Intra-D or Extra-D), type of training (Variable Irrel. or Constant Irrel.), and cue (color or form) relevant during shift. The analysis (Table 2) shows Extra-D Shift significantly more difficult than Intra-D Shift, and color more difficult than form. Training conditions (Variable Irrel. vs. Constant Irrel.) yield no significant main effect but do interact with type of cue relevant during shift. The low value of 2.0 for the C→F Constant condition is largely responsible for the T × C interaction. This effect may have come about because a strong cue (form) suddenly changing from constant value to variable attracts attention (see Discussion).

Learning curves.—The shape of individual learning curves for these *S*s can best be approximated with group data by means of backward learning curves (Hayes, 1953). In constructing backward learning curves, criterion day scores are combined for each *S* regardless of length of previous training. Scores are then averaged for criterion day minus one, criterion day minus two, etc. In Fig. 1 is shown the result of this computation of the three shift groups after Variable Irrel. training. The curves are extended to the left until the median *S* is dropped. Since most of the Rev. and Intra-D Shift groups met criterion the first day (7 out of 10 in both groups) their backward learning curves are for 1 day only. The median *S* in the Extra-D Shift condition took 3 days to meet criterion. It may be observed that the three curves are quite similar during the last five blocks of trials. Over the entire course of learning, performance of the Extra-D Shift group

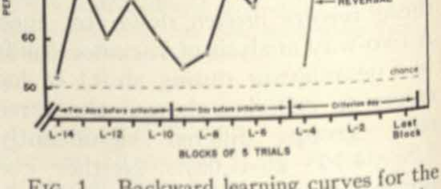


FIG. 1. Backward learning curves for the three shift groups after Variable Irrelevant training. (Curves are extended to the left until the median *S* drops out.)

may be described as ogival while curves for Rev. and Intra-D Shift are negatively accelerated.

As a test of the prediction from Wyckoff's model that if probability of observing relevant cues is decreased, the duration of chance performance will be increased, the first 10 trials of shift learning were analyzed. Figure 2 shows regular learning curves for the three shift conditions. It is evident that the Extra-D Shift group remains at chance throughout these 10 trials, while the other two groups show learning effects. A trend analysis showed a significant learning effect ($F=3.55$; $P<.01$) as well as a significant interaction between Shift and Trials ($F=4.18$; $P<.01$). Linear components of the shift curves differed ($F=5.16$; $P<.01$) as did also quadratic components ($F=6.77$; $P<.01$).

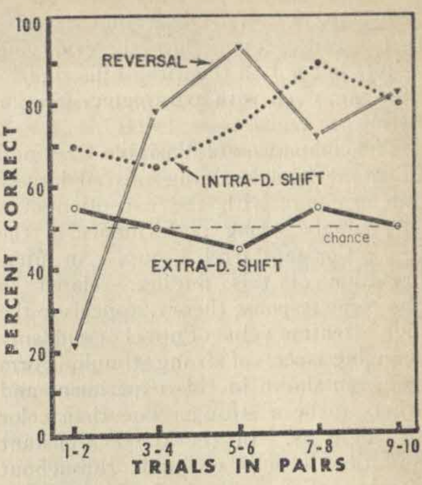


FIG. 2. Learning curves for the first 10 trials for the three shift groups after Variable Irrelevant training.

DISCUSSION

Performance on the shift tests following Variable Irrel. training yields evidence confirming predictions from "observing response" theory. Both the Rev. and Intra-D Shift groups show significant learning effects from the first trial indicating high probability of observing relevant cues. On the other hand, the Extra-D Shift group remained at chance for some trials as would be expected if probability of observing the relevant cues were low.

The superiority of Rev. over Extra-D Shift performance provides grounds for weighing the importance of transfer of a correct observing response. According to observing response theory, *S*s have the starting advantage in reversal of looking at the relevant cues, but the disadvantage of a strong tendency to pick up the wrong one. The *S*s under Extra-D Shift have an initial disadvantage of attending to irrelevant cues, but no corresponding disadvantage of a strong instrumental tendency to pick up the negative stimulus. If overall Rev. performance exceeds Extra-D Shift, therefore, it must be due to positive transfer of observing responses in the Rev. condition, and despite negative transfer of instrumental tendencies. The theory can account for this finding by assigning proper weights to the positive and negative transfer as these combine to determine performance, but the theory can also account for the opposite outcome by changing the weights. Consequently no rigorous test of theory will be provided by comparisons of Rev. and Extra-D Shift until a prior way of fixing parameters is found. The relative importance of observing and instrumental tendencies apparently depends upon the nature of the situation and *S*s, since Rev. has been found sometimes harder than Extra-D Shift and other times easier. Kendler and her associates (Kendler & Kendler, 1959; Kendler et al., 1960) failed to find Rev. faster than Extra-D Shift except for fast learners among 5 and 6 yr. olds, not for slow learners or younger children. For rats, Kelleher (1956) found Extra-D Shift faster than Rev. Reversal has been shown faster than Extra-D Shift for college students with a variety of experimental procedures (Buss, 1953, 1956; Harrow & Friedman, 1958; Kendler &

D'Amato, 1955). Observing response theory can accommodate these varying results but not all theories of discrimination can, even with parameters free for fitting.

The comparison of Variable Irrel. and Constant Irrel. training proved largely indeterminate with the exception of a surprisingly strong performance of the C→F Constant Irrel. group. An interpretation of this finding, related to observing response theory, appeals to the high attention value of novel or suddenly changing aspects of strong stimuli. Form has been shown in this experiment and others to be a stronger cue than color for these *S*s. In the C→F Constant condition form is constant throughout training, then suddenly begins to vary as the relevant dimension during the Extra-D Shift trials. The resulting combination of a relatively strong form cue and novelty apparently makes a highly discriminable dimension. Evidence that novelty or change is a readily discriminable aspect of stimuli for these *S*s has been presented in four previous experiments (House, Orlando, & Zeaman, 1957; House & Zeaman, 1958; Zeaman & House, 1962; Zeaman, House, & Orlando, 1958).

An alternative explanation of the shift effects, highly related to the present theory, has been invoked by Kendler et al. (Kendler & D'Amato, 1955; Kendler & Kendler, 1959; Kendler et al., 1960) to account for the superiority of reversals over extradimensional shift. This theory assumes mediation by implicit responses which add interoceptive cues to the external cues thereby making them more distinct. Kendler and Kendler (1959) have identified these symbolic mediating responses as verbal labels. Observing response theory is similar in that it assumes a mediating response is learned. The difference is that the observing response does not add cues but simply increases the probability that the relevant external cues will be perceived by the organism. The two theories are not incompatible, and it is entirely possible that both effects are present in the data reported here.

However, some of the effects which have been presented are not obviously related to the verbal label theory as presently stated. It is not clear how Intra-D Shift performance would be affected by learning labels which differentiate two specific stimuli. For example, the labels "red" and "green" should have little effect when colors black and yellow are relevant. To handle the Intra-D Shift results, verbal mediation theory might add the assumption that labeling is not a specific tenddency but rather a highly generalized response to an entire dimension of stimuli. Learning to label red and green might then be predicted to generalize to labeling black and yellow or any new colors when presented.

Even with this added assumption, verbal label theory does not share the advantage of observing response theory in predicting the observed shapes of the learning functions.

SUMMARY

Mentally defective children were trained on a discrimination problem with either color relevant and form irrelevant, or the opposite. One group had the irrelevant cue varied randomly during training; the other had a constant irrelevant cue. Following training, *S*s learned either a simple reversal, a new problem with stimuli from the same relevant dimension (intradimensional shift), or a new problem with stimuli from the irrelevant training dimension (extradimensional shift). Extradimensional shift was significantly more difficult than either reversal or intradimensional shift. Reversal and intradimensional shift performance did not differ.

The results were interpreted as consistent with observing response theory, but suggestive of changes for a verbal mediation theory.

REFERENCES

BUSS, A. H. Rigidity as a function of absolute and relational shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 153–156.

BUSS, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

HARROW, M., & FRIEDMAN, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–598.

HAYES, K. J. The backward curve: A method for the study of learning. *Psychol. Rev.*, 1953, **60**, 269–275.

HOUSE, B. J., ORLANDO, R., & ZEAMAN, D. Role of positive and negative cues in the discrimination learning of mental defectives. *Percept. mot. Skills*, 1957, **7**, 73–79.

HOUSE, B. J., & ZEAMAN, D. Reward and nonreward in the discrimination learning of imbeciles. *J. comp. physiol. Psychol.*, 1958, **51**, 614–618.

HOUSE, B. J., & ZEAMAN, D. Transfer of a discrimination from objects to patterns. *J. exp. Psychol.*, 1960, **59**, 298–302.

KELLEHER, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

KENDLER, T. S., KENDLER, H. H., & WELLS, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

KURTZ, K. H. Discrimination of complex stimuli. *J. exp. Psychol.*, 1955, **50**, 283–292.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.

WYCKOFF, L. B., JR. The role of observing responses in discrimination learning. Part I. *Psychol. Rev.*, 1952, **59**, 431–442.

WYCKOFF, L. B., JR. A mathematical model and an electronic model for learning. *Psychol. Rev.*, 1954, **61**, 89–97.

ZEAMAN, D., & HOUSE, B. J. Approach and avoidance in the discrimination learning of retardates. *Child Develpm.*, 1962, in press.

ZEAMAN, D., HOUSE, B. J., & ORLANDO, R. Use of special training conditions in visual discrimination learning with imbeciles. *Amer. J. ment. Defic.*, 1958, **63**, 453–459.

(Received February 9, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 452–457

THE CUE ADDITIVITY PRINCIPLE IN A RESTRICTED SOCIAL INTERACTION SITUATION[1]

JAMES M. RICHARDS, JR.

University of Utah

Recently, increasing attention has been paid to attempts to account for social phenomena in terms of three classic concepts of experimental psychology, namely stimulus, response, and reinforcement (Atkinson & Suppes, 1958, 1960; Burke, 1959; Suppes & Krasne, 1961). The heart of any such attempt is that when two people, A and B, are in a social interaction situation, and A makes some response to the total situation, that response is a stimulus, or cue, for B. When B then responds to this cue originating from A, his response in turn becomes a stimulus, or cue, for A. In other words, each person is a part of the environment of the other person and social cues are in no way qualitatively different from cues originating from other parts of the environment.

Within this general framework there are various ways in which social cues could combine with nonsocial cues, the simplest being through addition. Besides the advantage of simplicity, additive combination of cues offers at least two other major advantages: (*a*) that it makes possible the derivation of *quantitative* predictions through statement in terms of the learning theory developed by Restle (1955, 1957b), and (*b*) that the cue additivity principle has been shown to be a powerful explanatory concept for such diverse phenomena as place-and-response learning (Restle, 1957a), learning sets (Restle, 1958), and concept identification (Bourne & Restle, 1959). There have been, however, no specific tests of the cue additivity principle in social interaction. The present research, therefore, tests two specific hypotheses derived from the application of the general principle of additive combination of cues to social interaction. These are: (*a*) For each person in a two person social interaction situation cues originating from the other person combine additively with other nonsocial cues in the situation. (*b*) If a third person enters the social interaction, the situation is changed only in that the total number of cues is increased additively. More specifically, under the conditions of this experiment, where all persons involved are college age females, the number of cues added by each person is the same.

These hypotheses were stated in quantitative form through use of Restle's theory, which describes learning in terms of the single parameter θ. This parameter, θ, basically describes the rate of learning, and is defined as the proportion of the total cues in the situation which are "relevant" cues in the sense that they are "correlated" with reinforcement, or, in other words, that potentially *S* can use them as discriminatory cues to obtain reinforcement. The basic quantitative relationship tested is that the rate of learning, θ_{s+n}, for a group of *S*s for whom both social cues and nonsocial cues are relevant is equal to the sum of the rates of learning, $\theta_s + \theta_n$, of a group of *S*s for whom social cues are relevant and nonsocial cues are irrelevant and of a group of *S*s for whom nonsocial cues are relevant and social cues are irrelevant.

---

[1] Based on a dissertation submitted to the faculty of the University of Utah in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

METHOD

Subjects

The *S*s were 130 volunteer female elementary psychology students who received academic credit for participating. These *S*s were assigned to two conditions; 50 *S*s were randomly assigned, 10 to each of five parameter-estimating groups, and 80 *S*s were randomly assigned, 20 to each of four prediction-testing groups.

Apparatus

The apparatus consisted of an upright black panel to which two white geometric figures, a square and a triangle, were permanently attached. The purpose of this apparatus was to define the two responses available to *S*s in the experimental situation, and to provide the "nonsocial" cues. In addition to these nonsocial cues, there were various "background" cues (irrelevant presumably) originating from the experimental room, lighting conditions, internal states of *S*s, etc. It is assumed that these background cues are constant for all groups, and do not affect the equations.

Procedure

It was necessary to control the relationship between social cues and nonsocial cues for each *S*. This control was achieved through use of female stooges who responded in accordance with instructions from *E*. The *S*s were led to believe that the stooges were also *S*s and at no time, in so far as could be ascertained from observation, was any *S* aware that the other supposed *S*s in the experiment were, in fact, stooges. The same stooges were used throughout this experiment.

At the outset of the experimental procedure, each *S*, together with either none, one, or two stooges, was seated in front of the upright black panel and *E* was seated behind the panel. The following oral instructions were then given:

> This is a study of group processes. You see in front of you a panel on which are a triangle and a square. On each trial, each of you will say out loud either "triangle" or "square." In other words, you will choose first (*E* pointing to the person who was to choose first under that experimental condition) and say either "triangle" or "square," whichever you want. Then you (again pointing) will choose second and say either "triangle" or "square," whichever you want. Depending on what the two (three) of you choose, together you will be right or wrong, and I will tell you "right" or "wrong." Then you will choose again, and I will tell you "right" or "wrong," and so forth. Of course, what you are trying to do is to get to the point where you will be right all the time. The choices that are right one time may be, but are not necessarily, the choices that were right the time before. Are there any questions?

Any questions *S*s might have were answered, and then trials were run until *S*s reached a criterion of 10 successive correct responses.

The five parameter-estimating groups were run first, and θ values were computed from the mean number of errors on Trials 2 to criterion. These θ values were later combined to yield predicted mean number of errors in each of the four prediction-testing groups. These predicted means were then compared to the observed means of these four groups.

Parameter-Estimating Conditions

In four of the five parameter-estimating groups, only one stooge was present, and in the fifth no stooge was present. The five procedures are as follows:

Group 1.—The *S* chose first on each trial, and the stooge chose second. Nonsocial cues were relevant for *S*, and therefore the symbol for the learning rate parameter in this group is θ_n. The correct response for *S*s in this group was to choose a designated one of the two geometric figures and to ignore the responses of the stooge. For a randomly designated half of these *S*s, the figure chosen on Trial 1 was correct throughout, and for the other half the figure chosen was incorrect throughout. In this condition, the stooge chose in a simple alternation pattern, and, on Trial 1, chose the same figure chosen by *S* for half of the *S*s, and the opposite figure for the other half of the *S*s. This simple alternation pattern was chosen because it insured that social cues were uncorrelated with reinforcement, and therefore irrelevant, while at the same time insuring that the stooge's response on a given trial could be predicted perfectly from knowledge of her immediately preceding response.

Group 2.—The *S*s chose first on each trial and the stooge second, but with social cues relevant and nonsocial cues irrelevant. That is, reinforcement was given whenever *S* chose the same figure as the stooge regardless of what that figure was. The symbol for the learning rate parameter, therefore, is θ_s.

It was originally intended to have the stooge choose in a simple alternation pattern in this group too. Pretests involving 7 *S*s indicated, however, that this was not a satisfactory procedure, since *S*s tended, after a few trials, to fixate on a given figure and therefore never learn. Inquiry revealed that these *S*s were aware that reinforcement was given whenever they and the stooge chose the same figure, but had the hypothesis that the stooge was supposed to follow their choice. The following procedure was therefore used for this group. The stooge would choose in a simple alternation pattern for a block of 6 trials. After this block, if *S* chose the same figure for two successive trials, so would the stooge, thus providing information on the second of these trials disconfirming the hypothesis previously described. The stooge would then return to the simple alternation procedure for another block of 6 trials, etc. This pattern was repeated until *S*s were correct on 10 successive trials. On Trial 1, the stooge chose the same figure chosen by *S* for half of the *S*s, and the opposite figure for the other half of the *S*s.

Group 3.—The stooge chose first on each trial and the *S* chose second. The stooge chose in a 1212122121212112 pattern with the figure designated as 1 for *S* being randomly determined, but with the restriction that each figure be 1 for half of the *S*s. This pattern of responding was chosen for the stooge because it included all possible sequences of four responses for which the responses are divided 50-50 on each even numbered trial, thus making it possible to arrange reinforcement in such a way that social cues are irrelevant.

In this group, reinforcement was given when *S* ignored the choice of the stooge and chose a randomly designated one of the two geometric figures. That is, nonsocial cues were relevant.

Group 4.—The procedure was identical with that followed in Group 3, except that social cues were relevant and nonsocial cues irrelevant.

Group 5.—No stooge was present, and *S*s were reinforced for choosing a randomly designated one of the two geometric figures, with the restriction that each of the two figures was correct for half of the *S*s in this group. It was necessary to run this group in order to derive predictions for the three-person situation.

Prediction-Testing Conditions

Group 6.—One stooge was used, and *S* chose first on each trial. For a randomly designated half of the *S*s, the figure chosen on Trial 1 was correct, and for the other half incorrect. The stooge always chose the correct figure, so that both social cues and nonsocial cues were relevant. This made it possible to test predictions derived from the theoretical relationship $\theta_{s+n}=\theta_s+\theta_n$ where θ_n and θ_s were derived from the parameter-estimating groups in which *S* chose first.

Group 7.—The procedure was identical with that for Group 6 except that a second stooge was added to the situation, who chose after both *S* and the relevant first stooge had chosen. The responses of the second stooge were irrelevant, and were made in the 1212122121212112 pattern. Predictions for this group were derived from the assumption that the second stooge added the same number of cues to the two-person situation as the first stooge added to the situation in which no stooge was present.

Group 8.—One stooge was present who chose first and always chose a randomly designated one of the two geometric figures, with the restriction that each figure was chosen with half of the *S*s in this group. The *S*s were reinforced for choosing the same figure as the stooge chose, so that both social and nonsocial cues were relevant. Predictions for this group were based on the parameter-estimating groups in which the stooge chose first.

Group 9.—The procedure was identical to that for Group 8 except that a second stooge was added who chose before both the relevant first stooge and *S*. This second stooge chose in the 1212122121212112 pattern, and provided irrelevant cues. Again predictions were derived from the assumption that both stooges added the same number of cues.

In all of the above groups it proved necessary, because of scheduling considerations, to eliminate and replace all *S*s who did not learn to criterion within 150 trials. One *S* was replaced, therefore, in each of four groups: Group 1, Group 2, Group 5, and Group 7. Two *S*s were also eliminated because of errors by *E* in giving reinforcement.

RESULTS

The basic data for all groups in this experiment are the mean numbers of errors made from Trial 2 to the 10 criterion trials. Trial 1 is eliminated because on that trial *S*s had not received any information about which cues were correlated with reinforcement.

A value for θ corresponding to the mean for each of the five parameter-estimating groups was determined from a computing chart (Bourne & Restle, 1959) which presents θ as a function of mean number of errors to criterion. These means and their associated θ values are presented in Table 1.

The discrepancies between predicted and observed means for the prediction-testing groups were tested through use of the t test for significance of deviation from a point prediction. Bourne and Restle (1959) point out that this is the only convenient test currently available for such comparisons, although it is known to be, in some respects, inappropriate. Fortunately for the purposes of this study, the lack of appropriateness results in a conservative error, since a true theory would be rejected too often. Results of these tests are presented in Table 2.

TABLE 1

MEAN NUMBER OF ERRORS ON TRIALS 2 TO CRITERION AND ASSOCIATED θ VALUES FOR THE FIVE PARAMETER-ESTIMATING GROUPS

Group	Cues	Stooge	Mean	θ
1	Nonsocial, Relevant	None	2.80	.310
2	Nonsocial, Relevant	Chooses 1st	6.10	.171
3	Nonsocial, Relevant	*S* chooses 1st	11.60	.110
4	Social, Relevant	Chooses 1st	3.10	.281
5	Social, Relevant	*S* chooses 1st	5.40	.187

Although none of these differences was significant at the .05 level, there appears to be a reasonably large discrepancy between the predicted and observed values in Group 7. It appeared desirable, therefore, to try to determine whether this discrepancy was due to 1 or 2 *S*s or a definite tendency for the group as a whole to make more errors then predicted. As a partial check, therefore, a sign test was computed in which *S*s in this group making more errors than the predicted mean were given a plus and *S*s making fewer errors than the predicted mean were given a minus. In this group, 12 *S*s made more errors than the predicted mean and 8 *S*s made fewer, a result that is

TABLE 2

COMPARISON OF PREDICTED AND OBSERVED MEAN NUMBER OF ERRORS ON TRIALS 2 TO CRITERION FOR THE FOUR PREDICTION-TESTING GROUPS

Group	Condition	Predicted θ	Predicted Mean	Observed Mean	Observed SE_m	t
6	2-Person, Stooge 1st	.452	1.71	1.45	.2350	1.11
7	3-Person, Stooges 1st	.312	2.79	4.35	.7721	2.02
8	2-Person, *S* 1st	.297	3.00	3.30	.5264	.57
9	3-Person, *S* 1st	.181	5.05	5.15	1.1434	.09

Note.—For all t's, $df = 19$; none is significant at the .05 level.

TABLE 3

Comparison of Predicted and Observed Mean Numbers of Errors on Trials 2 to Criterion in the Four Prediction-Testing Groups Where Each Observed Mean is Compared to an Inappropriate Predicted Mean

Group From Which Predicted Mean is Taken	Group From Which Observed Mean is Taken	Predicted Mean	Observed Mean	Observed SE_m	t
3-Person, Stooges 1st	2-Person, Stooge 1st	2.79	1.45	.2350	5.70**
3-Person, *S* 1st	2-Person, *S* 1st	5.05	3.30	.5264	3.32*
2-Person, Stooge 1st	3-Person, Stooges 1st	1.71	4.35	.7721	3.42*
2-Person, *S* 1st	3-Person, *S* 1st	3.00	5.15	1.1434	1.88

$^{*}P < .01$.
$^{**}P < .001$.

obviously far short of significance at the .05 level.

Since none of the differences presented in Table 2 was significant at the .05 level, it also was felt that a check should be computed to determine whether, under the conditions of this experiment, this statistical procedure does typically lead to a rejection of the null hypothesis when *inappropriate* predictions are assessed. Accordingly, *t* tests were computed in which observed means for the three-person groups were tested against the predicted means for the two-person groups, and vice versa. Results are presented in Table 3.

Finally, in order to determine if the four prediction-testing groups differed reliably among themselves, a fixed constants model analysis of variance (McNemar, 1955) was computed with the rows representing order of choosing (i.e., whether *S* or the stooge chose first) and the columns representing the number of stooges present. In this analysis of variance, the dependent variable was the number of errors on Trials 2 to criterion for each *S*. Results are presented in Table 4.

TABLE 4

Analysis of Variance of Number of Errors on Trials 2 to Criterion for the Four Prediction-Testing Groups

Source	*df*	*MS*	*F*
Order of choosing	1	35.11	3.53
Number of stooges	1	112.81	11.33*
Interaction	1	5.52	.55
Within	76	9.96	—

$^{*}P < .01$.

Discussion

The major hypotheses in this experiment were that in social interaction, social cues combine additively with other cues in the situation and that under the conditions of this experiment, each person adds the same number of cues as every other person. Through use of Restle's learning theory, these hypotheses were stated in quantitative form, and the key test of their correctness is how accurately this theoretical formulation predicted the mean number of errors on Trials 2 to criterion for the four prediction-testing groups. The results presented in Table 2 indicate that in none of these groups was the observed mean number of errors significantly different from the theoretical prediction, in spite of the fact that the statistical analysis used would reject a true null hypothesis too often. On the other hand, the data presented in Table 3 indicate that when observed means are tested against inappropriate predictions, in two of these groups the differences are significant beyond the .01 level and in a third group, the difference is significant beyond the .001 level. This suggests that the failure to find significant discrepancies between predicted and observed values in Table 2 cannot be attributed to a lack of precision of the statistical test. Taken together, these results support the two experimental hypotheses and thus the operation of the cue additivity principle in social interaction.

It should be emphasized that in this experiment the "predictions" were, in fact, genuine predictions. The predicted values were computed prior to the collection of any data for groups used to test those predictions. This experiment, therefore, offered the advantage that the design made it possible to disconfirm the theory. The fact that such disconfirmation was not obtained, together with the fact that the predictions were quantitative rather than qualitative, provides additional support for both the operation of the cue additivity principle in social interaction and the power of Restle's theoretical formulation.

The results of the analysis of variance presented in Table 4 indicate that the only variable in the four prediction groups which had a significant effect was the number of stooges present. The fact that this variable had a significant effect tends to support the theory of Restle, since under the conditions of this experiment the second stooge was adding only irrelevant cues and thus should significantly increase the number of errors. It also provides evidence that the results of this study cannot be attributed to a lack of any real differences among the four prediction-testing groups. Order of choosing did not have a statistically significant effect. Since Restle's theory does not lead to predictions about the effect of this variable, this finding is not inconsistent with the theory. The fact that the interaction term in this analysis of variance was not statistically significant is further evidence that cues did combine additively.

SUMMARY

This study tested two hypotheses derived from the application of Restle's mathematical learning theory to social interaction: (*a*) For each person in a two-person social interaction situation, cues originating from the other person combine additively with other nonsocial cues in the situation. (*b*) If a third person is added to the social interaction, the situation is changed only in that the total number of cues is increased additively.

These hypotheses were tested in a group-learning situation, using groups consisting of 1 *S* and either 0, 1, or 2 stooges, with number of errors to criterion as the measure of learning. In no group was the observed mean number of errors significantly different from the predicted mean number of errors derived through use of Restle's theory. The data, therefore, support both hypotheses and provide evidence for the operation of the cue additivity principle in social interaction.

REFERENCES

ATKINSON, R. C., & SUPPES, P. An analysis of two-person game situations in terms of statistical learning theory. *J. exp. Psychol.*, 1958, **55**, 369–378.

ATKINSON, R. C., & SUPPES, P. *Markov learning models for multiperson interactions.* Stanford: Stanford Univer. Press, 1960.

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BURKE, C. J. Applications of linear models to person interactions. In W. K. Estes and R. R. Bush (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

RESTLE, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychol. Rev.*, 1957, **64**, 217–228. (a)

RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191. (b)

RESTLE, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, **65**, 77–91.

SUPPES, P., & KRASNE, F. Application of stimulus sampling theory to situations involving social pressure. *Psychol. Rev.*, 1961, **68**, 46–59.

(Received February 10, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 458–463

RESPONSE-MEDIATED GENERALIZATION WITH SIMPLE SKELETAL-MOTOR RESPONSES[1]

STANFORD H. SIMON[2]

Indiana University

Response-mediated generalization (RMG) may be said to occur when a response R_a, which has closely followed a second response R_b in one or more stimulus situations, follows R_b in a new stimulus situation to which only R_b has previously been conditioned. The mediated or generalized response would be R_a and the mediating response would be R_b.

While an eyeblink (e.g., Lumsdaine, 1939; Shipley, 1933, 1935) and verbal behavior (e.g., Birge, 1941; Cofer, Janis, & Rowell, 1943; Foley & Cofer, 1943; Wickens & Briggs, 1951) have been used as mediating responses, the writer knows of no experiments investigating the skeletal-motor response as a mediating response. There are several reasons for doing research in this area. In the first place, since the current learning theories have developed from data acquired predominantly from experimentation on skeletal-motor responses, such investigations would seem to be most readily related to present data on behavior and to present learning theories. Indeed, the prediction of RMG by present S-R theories would seem most applicable to the skeletal-motor response. In the second place, it would seem that some of the variables involved in RMG would be more accessible for intensive investigation if simple skeletal-motor responses were utilized. Such variables as the temporal relationship between the mediating and mediated responses, the duration of these two responses, and their magnitude could be investigated more readily because of the comparative ease with which simple skeletal-motor responses may be recorded.

The present experiment had three purposes. First, following the prediction of S-R theory (e.g., Hull, 1943), an attempt was made to demonstrate RMG with simple skeletal-motor responses serving as both the mediated and mediating responses. Secondly, RMG as a function of the number of pairings of the mediated response with the mediating response was investigated. This followed from the assumption in S-R theory that the chaining of responses takes place by the conditioning of a response to the response-produced stimuli (RPS) of the preceding response. The hypothesized RPSs are thought to function in conditioning in the same way as do external stimuli. Therefore, the strength of conditioning between the mediated response and the RPS of the mediating response would be expected to be a function of the number of times the responses have been paired. Since RMG is dependent upon this conditioning, RMG would be expected to be a function of the number of pairings of the mediated response with the mediating response.

---

[1] This report is based upon a thesis submitted in partial fulfillment of the requirements for the PhD degree at Indiana University. The author sincerely appreciates the assistance given him by W. K. Estes.

Microfilm of the complete manuscript, 66 pp., is available for $1.00 from University Microfilms; Ann Arbor, Michigan, Publication Number 5167. The thesis title is "Response-mediated generalization as a function of (*a*) the number of pairings of the mediating response with the mediated response, and (*b*) an external stimulus contingent on the mediating response."

[2] Now with the Psychology Service, Veterans Administration Hospital, Tomah, Wisconsin.

The third problem investigated was the effect of an external stimulus made contingent on the occurrence of the mediating response. The assumption involved here was that the strength of a learned response is a function of the number of stimuli (or strength of the stimuli) to which it is conditioned (Estes, 1950; Hull, 1943, p. 206). If the number of stimuli contingent upon the mediating response were increased, the mediated response could become conditioned to more stimuli; and if these were present when testing for generalization, more generalization would be expected than would otherwise be the case.

METHOD

Apparatus.—The *S* sat at one end of a table and faced the open end of an oblong box mounted on it. A panel board containing three lights located above the box was attached to the other end. Inside the box was a lever operated by *S* with his right hand. A telegraph key was operated by *S* with his left hand. The *S* wore a pair of earphones.

The box was approximately $1 \times .75 \times 3.25$ ft. Mounted perpendicular to the lever at *S*'s end was a handle covered with sponge rubber to reduce tactual cues of movement. (Visual and proprioceptive cues of movement were reduced by having *S* operate the lever in darkness and at arm's length.) The lever was 54 in. long. The handle displacement was 1 in.

The panel board was 25 in. wide and 26 in. high. Two inches from the top and 5 in. from either side were located two $\frac{3}{16}$-in. lights, the left one red and the right one green. A third light, $\frac{3}{8}$ in. in diameter, was located between the other two and was white. The lights were 4.5 ft. from *S*'s end of the table. The table with the apparatus was enclosed in a small lightproof cubicle.

Behind the cubicle and under the lever an Esterline-Angus recorder recorded key press, stimuli, and lever movement. The ratio of handle displacement to recording pen displacement was 1:4.5.

A noiseless switch operated by *E* in *E*'s booth, located behind and to the side of the cubicle, put either tone (S_t), red light (S_r), or green light (S_g) in circuit with a timer which presented the stimuli for .25 sec. at a rate of one every 3 sec. A second switch put the center white light in or out of circuit with the telegraph key. The tone was 800 cps and 15 db.

Both *S*'s and *E*'s cubicles were located in a semisoundproof room. While operating the apparatus *S* sat in darkness.

Subjects.—The *S*s, 120 students in introductory psychology at Indiana University, were under 25 yr. of age and right-handed.

Procedure.—The general experimental design fell into three phases. In Phase 1, *S* was instructed to press the key quickly whenever S_t or S_r occurred (S_g for control group). These were presented in random order for a total of 50 trials.

In Phase 2, *S* practiced operating the lever as well as the key. *All* *S*s were instructed to move lever left and back to center for S_r, and right and back to center for S_g. On any trial on which they were to do both, they were to always press the key first and *immediately afterward* move the lever. Only S_r and S_g were presented, randomly and for a total of 30 trials. No instructions regarding lever movement for S_t were given.

Phase 3 was the same as Phase 2 except S_t test trials were included on Trials 7, 10, 12, 16–25. There were 25 trials with S_r and S_g occurring randomly on non-S_t trials, 3 each occurring both before and after Trial 7. Here, RMG is defined as lever movement following key press to S_t in the same direction as lever movement previously associated with key press.

Group 1 followed the procedure described above and was divided into two subgroups: R and G. Subjects from R pressed the key for S_r and S_t; *S*s from G pressed the key for S_g and S_t. Since all *S*s moved the lever left for S_r and right for S_g, it was predicted R and G would differ in direction of lever movement following key press on S_t trials in Phase 3.

To investigate RMG as a function of the number of pairings of the mediated response with the mediating response, two additional groups were added. Group 2 was the same as Group 1 except *S*s did not use the key during Phase 2. Group 3 was the same as Group 1 except during Phase 2. Group 3 held the lever but did not operate it or the key. (No stimuli were presented during Phase 2 for this group; *E* stated he was adjusting the apparatus.) Groups 2 and 3, having less key press lever movement pairings, were predicted to show less RMG than Group 1.

To investigate RMG as a function of an additional stimulus occurring contiguously with the key press, three more groups were used. Group 4 was the same as Group 1 except that it had the white light in circuit with the key during Phases 2 and 3. Group 5 was the same as Group 4 except that the white light did *not* occur on the S_t trials of Phase 3. Group 6 was the same as Group 4 except that the white light occurred only on the S_t trials during Phase 3. Only in Group 4 was there both a chance for the lever movement to be associated with the white light and the white light present on S_t test trials in Phase 3. Therefore, only Group 4 was predicted to show more RMG than Group 1.

Groups 4, 5, and 6, having maximum training as did Group 1, were expected to show more generalization than did Groups 2 and 3.

Each of the six major groups had R and G subgroups with 10 *S*s randomly assigned to each of the 12 subgroups. Phase 1 was the same for all groups. The points of differences are summarized in Table 1.

Scoring.—Measures of direction and extent of lever movement on S_t trials in Phase 3 were taken. These were made in .04-in. units with the aid of a template. The first response of .04 in. or more occurring within ⅜ sec. after termination of the key press was recorded. (Because of the 1:4.5 ratio between the handle movement and the recorded movement, the criterion for a scorable response corresponds to .009 in. of actual movement of the handle by *S*.) For all *S*s, movements left were scored minus and movements right, plus. On the basis of previous pilot work it was decided beforehand to include only the first six S_t trials for this study.

TABLE 1

DIFFERENCES IN CONDITIONS BETWEEN GROUPS

Group	Number of Responses[a]		Number of Pairings[a]	White Light[b]
	Key	Lever		
1	52	36	36	Not present
2	37	36	6	Not present
3	37	6	6	Not present
4	52	36	36	Present
5	52	36	36	Present, except on S_t trials
6	52	36	36	Present only on S_t trials

[a] Total prior to first S_t trial in Phase 3.
[b] Contingent on key press responses in Phases 2 and 3.

"Amount of generalization" may be measured in two ways. Method 1 includes both the direction and magnitude of each response. Method 2 involves only the direction of the response. Both procedures involve the algebraic sum of the plus and minus scores over the first six S_t trials, and each has its own advantages and disadvantages. In this study both were used with the restriction that conclusions would be drawn only where differences were obtained at the 5% level of confidence by *both* procedures.

RESULTS

The *t* test was made on all comparisons, and these results as well as the scores by both scoring methods are given in Tables 2 through 4. When comparisons were made between the R and G subgroups of each major group, the scoring procedure outlined above was used, i.e., plus scores were assigned to movements to the right and minus scores to movements to the left. For comparisons between major groups, scores of the two subgroups were combined by giving plus scores to movements in the predicted direction (direction of generalization) and minus scores to movements in the opposite direction. This procedure was followed for both scoring methods.

The results will now be considered as they relate to the three problems raised by the experiment.

To test for RMG, Subgroups R and G of each of the six major groups were compared with each other. The means of the R groups were predicted to be in the minus direction (to the left) and those of the G subgroups in the plus direction (to the right).

Table 2 shows that the means of all subgroups, except R of Group 3, bear

TABLE 2

ANALYSIS ON SUBGROUPS R AND G USING FIRST SIX S_t TRIALS

Group	Scoring Method 1					Scoring Method 2				
	Mean		*SD*		*t*	Mean		*SD*		*t*
	R	G	R	G		R	G	R	G	
1	−9.56	11.20	12.46	15.57	3.12**	−1.62	1.90	1.49	2.55	3.59**
2	−11.20	7.30	18.61	13.61	2.40*	−.80	1.20	2.75	2.32	1.66
3	2.60	3.70	29.37	10.30	.11	.85	1.10	2.30	1.64	.27
4	−22.70	15.50	42.13	22.47	2.40*	−.70	2.90	2.61	2.02	3.27**
5	−40.20	20.14	49.90	41.29	2.80*	−1.65	1.78	2.67	2.24	2.96**
6	−8.70	14.34	9.50	9.50	2.51*	−1.80	2.54	1.66	1.80	5.29***

* $.05 > P > .01$.
** $.01 > P > .001$.
*** $P < .001$.

out this prediction with both scoring procedures. Statistical tests with both scoring procedures show a difference between R and G at the 5% level or more for all groups except Groups 2 and 3. Group 2 showed significance at the 5% level only with the first scoring procedure. Neither scoring procedure demonstrated a difference between R and G of Group 3.

In studying RMG as a function of the number of pairings of the mediated response with the mediating response we are concerned with differences between the minimally trained groups, Groups 2 and 3, and the maximally trained groups, Groups 1, 4, 5, and 6. (Groups 4, 5, and 6 also have the light as a variable.) Groups 2 and 3 were the only groups of the six major groups that did not show a significant difference between their two subgroups by both scoring methods. The question now remains whether or not the minimally trained groups differ significantly from the maximally trained groups.

TABLE 3

GROUP MEANS AND *SD*s USING FIRST SIX S_t TRIALS

Group	Scoring Method 1		Scoring Method 2	
	Mean	*SD*	Mean	*SD*
1	10.38	14.13	1.76	2.09
2	9.25	16.42	1.00	2.55
3	.55	22.23	.13	2.22
4	19.10	33.95	1.80	2.09
5	30.20	46.33	1.72	2.41
6	11.52	19.61	2.17	1.77

The analysis of variance over Groups 1, 2, and 3 was not significant with either Scoring Method 1 ($F = 1.71$, $df = 2/58$, $P > .05$) or Scoring Method 2 ($F = 2.42$, $df = 2/58$, $P > .05$). When the first scoring procedure is used the only group comparison that is significant at the

TABLE 4

RESULTS OF *t* TESTS BETWEEN GROUPS WITH BOTH SCORING METHODS

Group	*t*					
	1	2	3	4	5	6
1		.23	1.67	1.03	1.80	.21
2	1.82		1.37	1.14	1.86	.39
3	2.31*	1.13		1.99	2.52**	1.61
4	.06	1.92	2.39*		.84	.86
5	.05	.89	2.12*	.11		1.62
6	.63	1.64	3.14***	.60	.65	

Note.—Scoring Method 1 is above the diagonal and Scoring Method 2 is below the diagonal.
* $P = .05$.
** $P = .02$.
*** $P = .01$.

5% level is the difference between Groups 3 and 5 (Table 4). The second scoring procedure, which considered only direction of response, showed significant differences between Groups 3–1, 3–4, 3–5, and 3–6 (Table 4). All comparisons were in the predicted direction, i.e., the means of the maximally trained groups were larger than the means of the minimally trained groups, with both scoring procedures.

In studying RMG as a function of an external stimulus occurring contiguously with the mediating response we are concerned only with the groups that received maximum training. The comparisons made were between Groups 4–5, 4–6, 5–6, 1–4, 1–5, 1–6. No significant differences were found in any of these comparisons with either scoring method (Table 4).

DISCUSSION

Of the three hypotheses tested in this experiment, the evidence is clear for only two: the prediction that RMG can be obtained with simple skeletal-motor responses was verified; the prediction that a light made contingent on the occurrence of the mediating response would affect RMG significantly was not borne out. In general, the results seem to favor the third hypothesis, that RMG is a function of the number of pairings of the mediating response with the mediated response. However, since the results are not decisive, no definite conclusion was made.

If there is a real difference between Group 3 and the maximally trained groups, then it is difficult to say why the other minimally trained group, Group 2, did not tend to differ significantly from the maximally trained groups as did Group 3. It may be that while making the lever movement response in Phase 2 of the experiment, *S*s in Group 2 were also making key-press-like responses with their left hands even though they were not actually operating the telegraph key.

An important problem in this experiment is that of controlling the temporal relationship between the mediating and mediated responses. This may be related to the large variability between *S*s' scores in this experiment. While the instructions insured the completion of the key press response before the start of the lever movement response, it did not insure the same temporal relationship between these two responses for all *S*s.

Another problem involves the two scoring methods used in this experiment. The second scoring procedure differed from the first in that it resulted in less within-group variability and provided more significant differences. The first procedure, however, does not sacrifice information about the extent of movement as does the second procedure. If some of the problems involved in this experiment could be resolved with further research, e.g., the variability between *S*s in their temporal relationships between the mediated and mediating responses, the first scoring method might prove to be the more sensitive in demonstrating the effects of changes in the experimental variables.

SUMMARY

On the basis of current S-R theory, including assumptions concerning the functions of response-produced stimuli, it was predicted that response-mediated generalization (*a*) can occur with simple skeletal-motor responses, (*b*) will be greater for one group of *S*s than another if the first group has more pairings of the mediated response with the mediating response, and (*c*) will be greater for one group of *S*s than another if the first group has an additional stimulus occurring contiguously with the mediating response.

These predictions were tested in a simple motor learning situation with human *S*s. Pressing a telegraph key served as the mediating response and moving a lever served as the mediated response.

The experimental results support the first prediction but not the third; no definite conclusion is drawn regarding the second prediction.

REFERENCES

BIRGE, J. S. Verbal responses in transfer. Unpublished doctoral dissertation, Yale University, 1941.

COFER, C. N., JANIS, M. G., & ROWELL, M. M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32**, 266–269.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

FOLEY, J. P., JR., & COFER, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

LUMSDAINE, A. A. Conditioned eyelid responses as mediating generalized conditioned finger reaction. *Psychol. Bull.*, 1939, **36**, 650.

SHIPLEY, W. C. An apparent transfer of conditioning. *J. gen. Psychol.*, 1933, **8**, 382–391.

SHIPLEY, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12**, 337–357.

WICKENS, D. D., & BRIGGS, G. E. Mediated stimulus generalization as a factor in sensory pre-conditioning. *J. exp. Psychol.*, 1951, **42**, 197–201.

(Received February 23, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 464–467

ESCAPE PERFORMANCE AS A FUNCTION OF DELAY OF REINFORCEMENT [1]

HARRY FOWLER AND MILTON A. TRAPOLD

University of Pittsburgh *State University of Iowa*

The present investigation extends our program of studies on factors affecting escape performance (Bower, Fowler, & Trapold, 1959; Trapold & Fowler, 1960) to the delay-of-reinforcement variable. Specifically, the purpose of this study was to assess the effect of different delays of shock termination on the performance of rats running down a charged alley to a goal box that became uncharged at various delays following *S*'s entry into the goal. A wide range of delays in shock termination was investigated in order to provide a general picture of the function relating escape performance to the delay-of-reinforcement variable.

METHOD

Subjects.—The *S*s were 40 male, albino rats, approximately 180 days old at the start of the experiment. All *S*s had served previously as control animals in runway experiments wherein the conditions of food reinforcement imposed upon them were constant over training and uncorrelated with speeds. For a period of 2 weeks immediately prior to the start of the present study, *S*s were housed individually in the laboratory and were maintained on an ad lib. diet of Purina lab checkers and water.

Apparatus.—A straight alley, 68 in. long, 3 in. wide, and 4 in. high, was constructed of two L shaped strips of galvanized sheet metal, and was covered with a transparent Plexiglas top. One of the sheet metal strips formed one side and half of the floor of the apparatus; the other strip formed the other side and other half of the floor. A 1-in. gap separated the two strips at the floor.

The alley comprised three sections, a start box 8 in. long, an alley section 48 in. long, and a goal box 12 in. long. These sections were separated by upward acting guillotine doors. Raising the start-box door operated a microswitch which started a Standard Electric clock and closed a shock circuit to the entire apparatus. Breaking a photobeam 6 in. inside the alley stopped the first clock and started a second which, in turn, was stopped when *S* broke a second photobeam 2 in. inside the goal box. The breaking of the second photobeam also activated a Hunter timer which interrupted the shock circuit to the apparatus only after a specified delay following *S*'s entry into the goal box. The time data from the two clocks were recorded to the nearest .01 sec. and were converted to speed scores, i.e., (100/time). Hereafter, these scores will be referred to as starting and running speeds, respectively.

A single matched impedance, 60 cps, ac shock source provided shock to the apparatus. In this shock system, the galvanized metal strips forming the halves of the apparatus were connected across the output of a power transformer. In addition, a 0.25-megohm resistor was placed in series with *S*. The *S* received shock as long as its feet touched both of the metal strips forming the halves of the apparatus. The narrowness of the alley prevented *S* from running along only one side of the runway.

Procedure.—The *S* was placed in the start box with the start-box door closed and the goal-box door open. When *S* oriented toward the goal, the start-box door was opened and *S* could then run to the goal end of the alley. When *S* entered the goal box, the goal-box door was closed and *S* was detained in the goal box for 30 sec. or longer, as specified in the design section below. At the end of this time, *S* was removed from the goal box and immediately returned to the start box for the beginning of the next trial. In this manner, each *S* received a total of 28 highly massed trials.

Experimental design.—The 40 *S*s were assigned to eight groups of 5 each in such a way as to randomize past experimental history as well as all other extraneous factors. For all groups, intensity of shock in the apparatus was 240 v.; however, for different groups, this shock was terminated after different delays following *S*'s entry into the goal box and interruption of the final photobeam. Six of the eight groups of *S*s were assigned to one of the following delays in shock termination: 0, 1, 2, 4, 8, or 16 sec.

---

[1] This experiment was conducted at Yale University while H. Fowler held a predoctoral fellowship from the National Institutes of Mental Health, Bethesda, Maryland.

With termination of shock after the specified delay, *S*s of each of these six groups were detained in the goal box for an additional 30 sec. This procedure provided each of the six groups with the same length of time during which shock was absent at the goal, but confounded total length of time spent in the goal box (and consequently intertrial interval) with the delay variable. To account for this confounding, the two remaining groups of *S*s received different delays in shock termination at the goal, 8 and 16 sec., but with total time spent in the goal compartment held constant at 30 sec. for both groups. This feature of the design permitted possible extraneous effects resulting from differences in total goal-box time and intertrial interval to be assessed by a comparison of the performances of the main and alternate sets of 8- and 16-sec. delay groups.

It should be noted that the concomitant use of both the main and alternate delay conditions was additionally necessitated by the fact that the alternate condition, itself, confounds time during which the shock is absent at the goal with the delay variable. Emphasis was given to the main condition since it was anticipated that the confounding present with this condition would tend to offset rather than facilitate effects of the delay manipulation; the longer intertrial intervals associated with the longer delays should serve to increase speeds (e.g., Teichner, 1952), thereby reducing performance decrements attributable to delay of reinforcement.

RESULTS

Group mean performance curves over the entire course of training are presented for the starting and running speed measures in Fig. 1 and Fig. 2, respectively. To reduce the cluster of curves, the speed data for the alternate 8- and 16-sec. delay groups were not included in these figures. The data for these groups are presented below in relation to appropriate statistical comparisons.

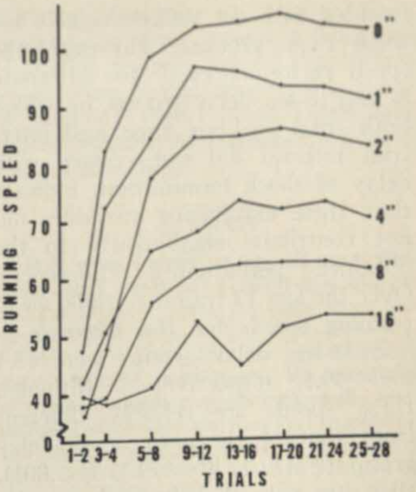


FIG. 1. Running speed (100/time in sec.) as a function of the delay (in sec.) of shock termination.

Running speeds.—Figure 1 shows that differences in mean running speed among the main delay groups emerged early in training, and increased progressively as the various groups approached different levels of asymptotic performance. Because of the generally orderly progression of the running speed curves over the course of training, statistical analysis of group mean differences was limited to the last 12 trials (17–28). The results of an analysis of variance of these data showed that the overall differences among the groups were highly reliable ($F = 6.0$, $df = 5/24$, $P < .001$). In addition, analysis by orthogonal polynomial comparisons of the trends within these data showed that mean running speed over Trials 17–28 was linearly related to the log of the delay values employed ($F = 30.0$, $df = 1/24$, $P < .001$); the residual variance after extraction of this linear-log component was negligible ($F < 1.0$).

As mentioned previously, total time spent in the goal box (and consequently, intertrial interval) are confounded with the delay variable for these main groups. However, the speed performances of the alternate 8- and 16-sec. delay groups, for which both total goal-box time and intertrial interval did not covary with delay of shock termination, indicate that these extraneous variables did not contribute significantly to the obtained performance differences. Over the last 12 training trials, mean running speeds for the *alternate* 8- and 16-sec. delay groups were 58.4 and 49.5, respectively. Although these speeds are reliably different from that of the zero or no-delay group ($F = 52.1$, $df = 1/12$, $P < .001$), they are only slightly and not significantly lower than those for the main 8- and 16-sec. delay groups ($F < 1.0$).

Starting speeds.—Figure 2 shows that the results obtained with the starting speed measure are generally the same as those for running speed. Differences in mean performance begin to emerge early in training and progressively increase until performance becomes asymptotic. Contrary to the running speed results, however, an analysis of variance of group mean starting speeds over the last 12 training trials showed that the overall differences among the delay groups were not significant ($F = 2.0$, $df = 5/24$, $P > .10$). Nevertheless, trend analysis of these data by orthogonal polynomial comparisons indicated that mean starting speed over these trials, like running speed, was linearly related to the log of the delay values employed ($F = 9.7$, $df = 1/24$, $P < .005$); again, the residual variance after extraction of this linear-log component was negligible ($F < 1.0$).

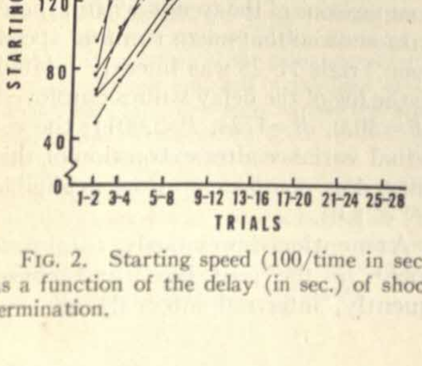


FIG. 2. Starting speed (100/time in sec.) as a function of the delay (in sec.) of shock termination.

The pattern of results obtained for the alternate 8- and 16-sec. delay groups with the starting speed measure was the same as that obtained with running speed. Mean starting speeds for these groups over the last 12 training trials were 165.5 and 148.9, respectively. Again, these speeds are reliably different from that of the no-delay group ($F = 15.1$, $df = 1/12$, $P < .005$), but only slightly and not significantly lower than those for the main 8- and 16-sec. delay groups ($F < 1.0$). Hence, these data also indicate that the obtained differences are not attributable to the covariation of intertrial interval or total goal-box time with delay of shock termination.

DISCUSSION

In total, the results show that for rats running down a charged alley speed is significantly decreased when reinforcement, i.e., shock termination, is delayed following the behavior which is instrumental to it. Further, this decrement in performance is a decreasing, negatively accelerated function of the delay interval, at least over the range of values employed in this study. In terms of the two response measures used, these effects were quite pronounced with running speeds and somewhat less so with starting speeds.

In the context of the behavior theory

developed by Spence (1956) for instrumental appetitive conditioning, the effects of delayed shock termination can be accounted for by the inhibitory intervening variable, I_t, the basis for this inhibition being competing responses which are conditioned during the delay period. In the present study, *S*s were observed to exhibit, following their entry into the goal box and in so far as the length of the delay period permitted, responses such as turning about, cringing, pushing against the goal-box cover, etc. From the competing response theory of I_t, these responses, which become conditioned to the cues at the goal, and hence anticipatory by way of the similarity of the cues in the alley, would tend to interfere with running and thereby reduce speed. In the same vein, the differential effect of the delay manipulation on starting and running speeds can be attributed to a decrease in the similarity of alley and goal-box cues, the greater the distance from the goal. Finally, these inhibitory effects should be more pronounced with longer delays, since the opportunity for the interfering responses to occur and become conditioned will be greater, the greater the length of the delay period.

With the previous studies in this series, additional parallels between appetitive and escape conditioning were demonstrated in the similarity of operation of both the drive and magnitude-of-reinforcement variables. For rats running down a charged alley to a goal box which is either completely or partially uncharged, escape performance is positively related to both shock intensity in the alley (Trapold & Fowler, 1960) and magnitude of shock reduction at the goal, with shifts in the magnitude of goal-shock reduction leading to rapid and appropriate shifts in level of performance (Bower et al., 1959). Taken together, the total findings of this series of studies strongly suggest that escape conditioning can be interpreted in terms of behavior theory formulations now offered primarily for instrumental appetitive conditioning.

SUMMARY

Rats were trained to run from an alley where they received a continuous electric shock of 240 v. to a goal box where shock was terminated at either 0, 1, 2, 4, 8, or 16 sec. following *S*'s entry into the goal compartment. Escape performance was measured in terms of response speeds both at the start and alley sections of the runway. In general, the results showed that speeds were slower the greater the delay of shock termination at the goal. This effect was quite pronounced for speeds in the alley section of the runway, but considerably less so for speeds at the start section. These results are analogous to those obtained in runway situations where delayed appetitive rewards have been employed and, as such, they indicate, along with those findings previously reported, that similar behavioral laws underlie both appetitive and escape conditioning.

REFERENCES

BOWER, G. H., FOWLER, H., & TRAPOLD, M. A. Escape learning as a function of amount of shock reduction. *J. exp. Psychol.*, 1959, **58**, 482–484.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

TEICHNER, W. H. Experimental extinction as a function of the intertrial intervals during conditioning and extinction. *J. exp. Psychol.*, 1952, **44**, 170–178.

TRAPOLD, M. A., & FOWLER, H. Instrumental escape performance as a function of the intensity of noxious stimulation. *J. exp. Psychol.*, 1960, **60**, 323–326.

(Received February 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 468–471

RESPONSE STRENGTH AS A FUNCTION OF DRIVE LEVEL AND PRE- AND POSTSHIFT INCENTIVE MAGNITUDE[1]

DAVID EHRENFREUND AND PIETRO BADIA

Adelphi College

About ten years ago Hull (1951), largely on the basis of Crespi's (1942, 1944) reports, raised the status and importance of the incentive magnitude variable by removing it as a determiner of H and introducing a separate incentive motivational intervening variable, K, which is designed to handle both differences and shifts in incentive magnitude. The manner in which differences in K combine with other intervening variables particularly D has not yet received enough empirical support to justify, with any large degree of confidence, specifying an exact function. Hull proposed a multiplicative function but, as Spence (1956) has pointed out, this was purely a working hypothesis since there was no evidence.

Although Spence (1956) suggested a linear function (Ramond, 1954), he (Spence, 1960) later suggested that . . . "it is possible that this will have to be changed to some nonlinear function" (p. 96). To date, however, runway studies, at least, have tended to support the additive function (Reynolds & Pavlik, 1960; Weiss, 1960).

The manner in which D combines with shifts in incentive magnitude would also seem to be important theoretically although there have been no reported empirical studies.

The present study involved a 2 × 2 factorial design involving two levels each of drive and reward. After asymptotic speeds were presumed to be reached, the magnitude of incentive of each group was shifted to that of the other.

METHOD

Subjects.—The *S*s were 20 Sprague-Dawley strain white rats, approximately 100 days old, equally divided among four groups: high drive-low reward, high drive-high reward, low drive-low reward, and low drive-high reward. One *S* in the last group had to be discarded, reducing the *N* for this group to 4.

Apparatus.—The sides and bottom of the runway apparatus were constructed of cloudy plastic. The top was clear plastic covered with two thicknesses of fine wire screen. The runway was 5 ft. long, and 2½ in. wide, and 4 in. high. It was sandwiched between two large opaque boards which measured 6×2½ ft. The space between these boards was completely enclosed around the outside border. A 1 in. × 5 ft. slit was made in the bottom board beneath the center of the floor of the maze. With this arrangement, filtered light beams were directed from above the maze to photoelectric cells below the maze. In the present study speed was recorded in the middle 2-ft. section of the maze with a Standard Electric timer to .01 sec. The maze was illuminated by four 7½-w. bulbs along each side. These were cut down by a variac to 40% and the cloudy plastic sides diffused the lights evenly in the maze. Guillotine type cloudy plastic doors, 15 in. from each end of the maze created a start box and goal box. Just in front of this door in the start box there was another horizontally sliding door of clear plastic. Thus walls, floor, and ceiling (except for slits for the doors) were continuous, homogenous, and unbroken from one end of the runway to the other.

Drive and reward.—Drive was operationally defined in terms of a percentage of *S*'s weight under conditions which maintained this weight continuously thus precluding long periods of deprivation. During the experiment all *S*s lived in a weight control apparatus, the details of which are described in a previous paper (Ehrenfreund, 1960). Es-

---

[1] Supported by a grant from the National Science Foundation.

TABLE 1

INDIVIDUAL GROUP ANALYSIS OF ASYMPTOTIC SPEED SCORES

Reward	Body Weight			
	95%*		85%	
	Mean	*SD*	Mean	*SD*
45 mg.**	10.26	.82	13.92	1.34
260 mg.	13.86	1.27	15.15	1.04

* $P < .005$ for low-drive groups.
** $P < .005$ for low-reward groups.

sentially each *S* is maintained within 1% of a specified weight. Loss of weight in the order of up to 2 gm. results in the automatic delivery of a 97-mg. Noyes tablet every 12 sec. until the specified weight is regained. Water is always available. Under such a schedule *S*s receive small amounts of food aperiodically around the clock. For the purposes of this study the high-drive *S*s were maintained at 85% of ad lib. or starting weight and the low-drive *S*s at 95% of ad lib. weight.

On each trial the low-reward group received a 45-mg. Noyes tablet. The high-reward *S*s received a 260-mg. tablet.

Procedure.—When *S* faced the vertical door of the start box it was raised and 2 sec. later the horizontally sliding clear plastic door was opened. Since under the present drive maintenance schedule *S*s are always at the stated drive level, data could be collected several times a day. All *S*s received 90 acquisition trials under the original drive-reward conditions. On Trial 90, each *S* in a high-reward group (260 mg.) was shifted to a low reward (45 mg.), and vice versa for 25 post-shift trials. Trials were run in blocks of 5 with 15 to 20 sec. between trials and from ½ to 1 hr. between 5-trial blocks. Between blocks of trials, *S*s were returned to the weight control apparatus. All of the 115 trials were run over a 4-day period; 20 on Day 1, 30 on Days 2 and 3, and 40 on Day 4. The median of 5-trial blocks for each *S* was converted to its reciprocal. The means of the scores thus converted form the bases for the tables, graphs, and statistical analyses.

RESULTS AND DISCUSSION

Prereward shift considerations.—The last two blocks of trials before the shift in reward magnitude were assumed to reflect asymptotic values.

An analysis of variance of these values revealed that the main effects of Drive and Reward were both significant beyond the .001 level, the *F* values being 19.85 and 18.90, respectively. The *F* value for interaction was 4.52 which is significant at the .05 level. Table 1 supplies the analysis for the individual group means. Note that the difference between the two low-drive groups (or low reward) is greater than that between the high-drive groups (or high reward). A simple plot of these asymptotic means reveals that the departure from linearity is converging as demanded by exponential addition rather than diverging, as demanded by a multiplicative function.

In order to facilitate comparison of the acquisition curves, the four groups are paired in different combinations in Fig. 1 and 2. The delayed separation of the curves for the two high-drive groups, in contrast to those of the two low-drive groups, lends further support to a negatively accelerated function. Under conditions of either high drive or high reward, the differential increments of combined *D* and *K* would be small.

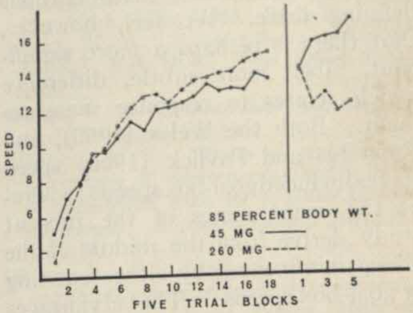


FIG. 1. High-drive groups rewarded with either 45 mg. or 260 mg. (To the right of the vertical line are shown the effects of changing the reward of each group to that of the other group.)

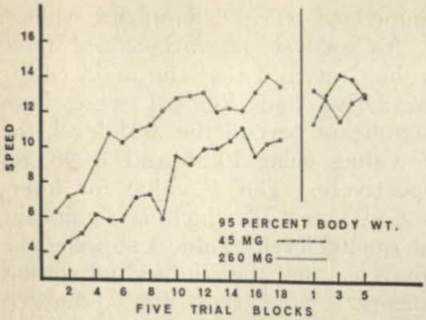


FIG. 2. Low-drive groups rewarded with either 45 mg. or 260 mg. (To the right of the vertical line are shown the effects of changing the reward of each group to that of the other group.)

On the assumption that habit strength multiplies with this combination, $E = H \times (D + K)$, the separation of the groups would not be manifest until H had developed some appreciable magnitude. These data are not in agreement with two recent studies (Reynolds & Pavlick, 1960; Weiss, 1960) which failed to find an interaction between drive and reward and therefore supported a linear or simple additive function.[2] Since both of these studies employed hours of deprivation as contrasted with the weight control method of the present study, the discrepancy might be attributed to the different operations defining drive. We feel, however, that there is perhaps a more significant, albeit more subtle, difference which relates to response measurement. Both the Weiss (1960), and Reynolds and Pavlick (1960) speed scores included goal-box speeds; whereas the speed scores of the present study derive from the middle of the alley, uninfluenced by either starting or goal-box speeds. The advantages of delineating at least these three response speeds (i.e., starting, running, goal) has recently been demonstrated by Goodrich (1959) and Wagner (1961) whose data are particularly convincing in reference to goal-box versus running speeds.

The effect of shift in incentive magnitude.—The immediate effect of the shift in magnitude was an increase in general excitement for practically all *S*s, as revealed in increased difficulty of handling, resistance to being placed in the start box, and, in some cases, a delay in the consummatory response even with those *S*s shifted upward. These behaviors were also reported by Crespi (1944) who, however, only mentioned them in connection with *S*s shifted downward. Not only did we observe them for *S*s shifted upward (although to a lesser extent) but more significantly, in terms of the analysis of the present data; they were clearly more obvious in the high-drive *S*s.

The right sides of Fig. 1 and 2 show the effect of shifting reward from 260 mg. to 45 or from 45 mg. to 260. Our main interest is centered on the fact that both the elation and depression effects obtain under high drive ($P < .05$), but neither obtains under low drive. Assuming this to be a genuine phenomenon, the problem remains as to how to account for it within the theory. The evidence to date is much too sparse to justify a detailed theoretical analysis. Nevertheless, existing theoretical concepts elaborated in other experimental contexts, coupled with qualitative observations of *S*s' behavior mentioned above, suggest the importance of the role of excitement, or the fractional anticipatory emotional response r_e as developed by Spence (1956, 1960) or r_f as developed by Amsel (1958).

As the data now stand, no new assumptions other than those already developed concerning the role of r_e are required. Occurring originally in the goal box, r_e

---

[2] It might be pointed out, however, that the data of the present study are exactly in line with the type of additive hypothesis that Spence (1953, p. 20) proposed in the Kentucky Symposium.

would generalize to the highly similar runway. Also r_e tends to elicit through its response produced cues $(r_e \rightarrow s_e)$, overt responses, some of which may be incompatible with the running response. Thus the occurrence of r_e in combination with the existing drive level will increase or decrease response strength to the extent that experimental manipulations tend to facilitate or preclude the development of such competing responses of varying strengths. In the present situation it is hypothesized that such incompatible responses are much more likely to occur under conditions of reduction in incentive magnitude. This will be recognized as essentially the same theoretical analysis made by Spence (1960) for extinction phenomena which are here regarded as the limiting case of incentive magnitude reduction. Accordingly, the postshift response decrements and increments are seen as functions of drive level and the difference and direction between pre- and postshift reward magnitudes. In these terms, the so called elation and depression effects would become merely points on a curve and thus disappear as independent phenomena.

SUMMARY

In an attempt to ascertain (*a*) the function describing how *D* and *K* combine and (*b*) the effect of incentive magnitude shifts as a function of *D*, speed measures were obtained on four groups of rats involving two drive levels and two reward levels. After 90 trials on the original reward level, the reward of each group was shifted to that of the other for 25 additional trials. Statistical analyses revealed a significant interaction ($P = .05$) between *D* and *K*, indicative, not of a multiplicative function but of a negatively accelerated one such as exponential addition. Both main effects were highly significant but *t* tests were significant only between the two low-drive groups and between the two low-reward groups. After the shift in incentive magnitude, changes in running speed occurred in the expected direction but the differences were much greater under high drive than under low drive. It was hypothesized that such differences (i.e., Crespi effect) were an increasing function of *D* and the extent of the shift in incentive magnitude. The shift data were interpreted in terms of the hypothetical anticipatory emotional response mechanism $(r_e \rightarrow s_e)$.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

CRESPI, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

EHRENFREUND, D. The motivational effect of a continuous weight loss schedule. *Psychol. Rep.*, 1960, **6**, 339–345.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

RAMOND, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, **47**, 248–250.

REYNOLDS, W. F., & PAVLIK, W. B. Running speed as a function of deprivation period and reward magnitude. *J. comp. physiol. Psychol.*, 1960, **53**, 615–618.

SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In, *Learning theory, personality theory, and clinical research: The Kentucky symposium.* New York: Wiley, 1953. Pp. 1–21.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* N. J.: Prentice Hall, 1960.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WEISS, R. F. Deprivation and reward magnitude effects on speed throughout the goal gradiant. *J. exp. Psychol.*, 1960, **60**, 384–390.

(Received February 27, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 472–479

THE INDUCTION OF NONVERIDICAL SLANT AND THE PERCEPTION OF SHAPE[1]

WILLIAM EPSTEIN, HELEN BONTRAGER, AND JOHN PARK

University of Kansas

Koffka's (1935, pp. 224–235) explanation of shape-constancy is based on the assumption of an invariant linkage between slant and shape. Koffka's hypothesis has been formulated in the following way by Beck and Gibson (1955): "A retinal projection of a given form determines a unique relation of apparent shape to apparent slant" (p. 126). The invariance hypothesis requires that perceived shape vary as a strict function of variations in perceived slant (Langdon, 1951, 1953; Nelson & Bartley, 1956; Stavrianos, 1945).

In order to test this hypothesis Beck and Gibson (1955) induced errors in the apparent slant of a target whose shape was to be judged. A triangular target cut from untextured, white cardboard was mounted on a regularly textured, vertical background, and was slanted outward from the base at an angle of 45° from the background. The target was viewed monocularly with a motionless head. Under these conditions, in the absence of binocular disparity, motion perspective, and texture gradients, the slant of the target was misperceived. The triangle assumed the slant of the background, i.e., it appeared to be perpendicular to the line of sight instead of slanted 45°. Along with this standard target, two comparison triangles placed flat on the background were exhibited. One represented a frontal parallel projection of the standard, and the other had the same physical dimensions as the standard. All *S*s judged the standard triangle to be equal in shape to the projective comparison. This match agrees perfectly with the requirements of the invariance hypothesis. When the same judgments were made with unrestricted binocular vision, 77% of the *S*s selected the objectively equal comparison triangle. This result also is in the direction required by the invariance hypothesis.[2] Since unrestricted observation eliminates the slant-inducing effect of the background, accurate perception of slant was restored; and, with it, veridical perception of shape was also recovered.

The general purpose of the present study was to extend and clarify Beck and Gibson's (1955) findings. The following were the major extensions and modifications: (*a*) The slant-induction effect was investigated when the background also was slanted from the frontal parallel plane. (*b*) A technique was employed which permitted continuous variation of the comparison stimulus. This made possible a more exact test of the invariance hypothesis. (*c*) A more accurate measure of the apparent slant of the target was obtained. These data are necessary for an analysis of the slant-shape relationship. (*d*) The influence of three different instructional sets on the judgment of shape was investigated. Beck and Gibson (1955) do not inform us concerning this variable, and the evidence from other studies (Gottheil & Bitterman, 1951; Klimpfinger, 1933) is fragmentary.

METHOD

Apparatus

The main apparatus was a rectangular light-tight tunnel 7 ft. in length with walls

---

[1] This investigation was supported by grants to the first author from the National Institute of Mental Health of the United States Public Health Service (M-4153) and the General Research Fund of the University of Kansas. We wish to acknowledge the assistance of Raymond Engstrand who, together with the second author, collected the data and performed the statistical computations.

[2] The invariance hypothesis actually requires a 100% shift to the objectively equal comparison.

20 in. wide. The interior of the tunnel was painted flat black. In the front wall of the tunnel there was an aperture which could be adjusted for monocular or binocular vision. A head-clamp chin-rest arrangement placed in front of the aperture kept *S*'s head motionless when this was required. The two clamps restrained all but the most determined head movements, and head movements are considered highly unlikely under the conditions of the experiment. A reduction tube was inserted in the front wall which restricted the monocular field to the target and its immediate background. On the inside of the front wall was a circular fluorescent lamp (32 w., standard cool white color) which provided the only illumination in the tunnel. This lighting arrangement made invisible the shadow cast by the slanted standard. At a distance of 6 ft. from the front wall a false back wall (panel) was inserted. This panel was covered with a black and white checkerboard cloth composed of 1 × 1 in. squares. The panel could be adjusted from outside the tunnel to three degrees of slant: perpendicular to the line of sight (vertical), 20° from the perpendicular tilted away from *S* (20° A), and 20° from the perpendicular tilted toward *S* (20° T).

The standard triangle was mounted in the objective center of the panel and directly in *S*'s line of regard, and had a height of 5 in. and a base of 4 in. The adjustable comparison triangle was mounted flat on the background, 4 in. above the apex of the standard. The height and base of the comparison could be varied continuously by manipulating a set of levers located on the exterior of the roof of the tunnel. The design of this comparison stimulus is similar in essentials to the apparatus described by Gottheil and Bitterman (1951, p. 407). The comparison and standard triangles were cut from the same white cardboard of imperceptible texture.

The comparison stimulus for slant was a circular disc of white cardboard mounted on a horizontal axis. The disc could be rotated on its axis, and its angle of rotation could be read directly off a protractor. The disc was located directly in *S*'s line of sight when he turned 90° into the designated viewing position, and it was presented outside the tunnel in a normally illuminated, unrestricted field.

Procedure

Stimulus conditions.—The standard was judged twice by each *S* (one ascending and one descending trial) for each of three background slants. When the background was vertical or 20° T, the standard triangle was at a 45° slant from the background outward from the base. When the background was slanted 20° A, the standard was slanted 45° outward from the apex.

Conditions of observation.—Each *S* observed the standard either monocularly or binocularly.

Instructional conditions.—Three sets of instructions were used which were intended to induce different attitudes of observation. These attitudes are usually designated: the phenomenal, objective, and analytic. Each *S* served under one instructional condition only. The instructions were as follows:

> (Instructions for the phenomenal attitude) When you look into this box you will see a standard triangular target on a checkerboard background. I am interested in learning how you perceive the shape of this target. Right above the standard triangle there is a second triangle. The base and altitude of this second triangle can be varied. Your task will be to instruct me to adjust the shape of the variable triangle so that it appears to be the same shape as the standard below it. I would like you to suspend all mental judgments and give me a match which reflects your immediate perceptual impression. Don't try to figure out a good match. I want a report of your immediate perceptual impression even if you feel that the match you make would not agree with the objective physical dimensions of the target.
>
> (Instructions for the objective attitude) I am interested in learning how well you can reproduce the actual physical shape of this target. . . . This means that in the ideal case when you have completed your judgment I should be able to take the match you have made, lay it over the standard, and find that it corresponds perfectly in all dimensions. Please remember that you are to reproduce the actual physical dimensions of the target even if the match you make doesn't *look* equal in shape to you. To make this clear suppose you were looking at a man far in the distance. He would *look* very small but if you were asked to reproduce the actual size of the man you probably would be fairly accurate. This is what I wish you to do here. Reproduce the actual physical shape of the target.
>
> (Instructions for the analytic attitude) I am interested in learning how well you can reproduce the retinal shape of this target. An illustration may help make this

> clear. If I put a coin on the table and you look directly down upon the coin, it will project a circular retinal image. However, if you stand back along the edge of the table and look at the coin from this new position, the coin will produce an elliptical retinal image. It is this retinal shape in which I am interested. . . . Your task will be to instruct me to adjust the shape of this variable triangle so that it is identical in shape with the retinal shape of the standard triangle. Remember I am interested in the retinal shape. It is not important to me whether or not the two triangles look identical in shape or are actually equal in shape when you have completed your judgment.

The shape-matching procedure may now be summarized. Each *S* was instructed to make two shape-judgments of the standard for each of three background slants. The *S* made the judgments by instructing *E* to adjust the shape of the comparison triangle until a satisfactory match was attained. Each *S* performed this task under one observational condition, binocular or monocular, and under one instructional set, phenomenal, objective, or analytic.

When the shape-judgments were concluded, the standard was presented again, and *S* judged the slant of the standard once for each slant of the background. All *S*s were given the following instructions:

> Now I would like you to judge the slant of the target triangle. By "slant" I am referring to whether the triangle appears to stand straight up and down in front of you, or whether it appears to lean either toward or away from you. This is how your slant judgments will be made. First, you will look at the triangle and determine if it is slanted and, if so, how much. Outside of the box I will show you a circular disc whose slant may be varied. Your task is to adjust the slant of this disc to the same slant as the target. After you have adjusted the slant of the disc you will be given one more look into the box to check your judgment.

The experiment was concluded with an interview which sought to obtain information about three aspects of the experiment: (*a*) the effectiveness and comprehensibility of the instructions; (*b*) the deliberate utilization of perceived slant during the judgment of shape; and (*c*) the relation between the judgments of slant obtained *after* the shape-judgments were completed and the perceived slant *during* the shape-judgments.[3]

Subjects

The *S*s were 84 undergraduates who served in the experiment as a course requirement. They were assigned in order of their appearance to one of six groups (2 conditions of observation × 3 attitudes of observation). Thus, each of the six experimental groups contained 14 *S*s, about equally divided among men and women. None of the *S*s knew anything about the invariance hypothesis, or the questions under investigation.

RESULTS

The main results of the experiment are presented in Table 1. In all of the tables the shape data are represented as height-base (h/b) ratios.

Slant-induction.—An inspection of the mean slant judgments for monocular observation recorded in Table 1 reveals that the slant-inducing effect of the background reported by Beck and Gibson (1955) was obtained for all three background slants. As an illustration consider the fourth row of figures in Table 1. When the background was 20° A and the standard was objectively slanted 65° A, the standard was judged to be slanted 17.21° A. When the background was vertical and the standard 45° T, the standard was judged to be slanted 0.83° T. In the case of the 20° T background slant, an exaggerated slant-induction effect appears to have occurred. While the standard was objectively slanted 65° T, it was perceived to be at a slant of 6.59° T, i.e., slanted 13.41° less than the background.

The comparable data for binocular vision show the expected diminution of the slant-induction effect when the

---

[3] The use of successive rather than simultaneous slant-shape judgments can be questioned. We sought to eliminate some of the ambiguity during the interview.

TABLE 1

Mean h/b Ratios of Comparison Matches and Mean Slant Judgments of Standard

Observation Attitude	Slant of the Background											
	20° A				Vertical				20° T			
	Comparison		Standard		Comparison		Standard		Comparison		Standard	
	Mean h/b	*SD*	Mean Slant	*SD*	Mean h/b	*SD*	Mean Slant	*SD*	Mean h/b	*SD*	Mean Slant	*SD*
	Monocular Vision											
Phenomenal	.370	.043	10.57° A	11.40°	.573	.053	0.80° T	3.37°	.386	.063	4.06° T	7.58°
Objective	.397	.018	17.30° A	12.00°	.556	.043	0.85° T	3.58°	.386	.044	10.21° T	6.48°
Analytic	.387	.097	23.57° A	13.96°	.530	.090	1.15° A	4.75°	.392	.091	7.64° T	7.54°
All attitudes	.384	.127	17.21° A	13.03°	.553	.180	0.83° T	4.43°	.388	.128	6.59° T	7.49°
	Binocular Vision											
Phenomenal	.547	.177	55.72° A	9.57°	.649	.077	20.57° T	8.89°	.584	.117	25.00° T	10.63°
Objective	.706	.102	52.72° A	14.07°	.710	.095	17.57° T	9.45°	.666	.129	35.28° T	12.53°
Analytic	.612	.153	56.65° A	7.91°	.631	.079	20.35° T	6.74°	.507	.134	29.35° T	11.27°
All attitudes	.625	.202	55.03° A	11.40°	.696	.213	19.49° T	7.28°	.585	.192	29.87° T	12.36°
Objective h/b (St.)	.800				.800				.800			
Proj. h/b (St.)	.353				.585				.324			
Objective slant (St.)[a]	65° A				45° T				65° T			

[a] Measured in terms of the deviation from the vertical away from (A) or toward (T) *S*.

cues for slant were restored. The case of the 20° A background illustrates this diminution. When viewed monocularly, the target appeared to be at a slant of 17.21° A—a deviation of 47.79° from its objective slant, but only 2.79° from the slant of the background. However, when viewed binocularly, the target was judged to be at a slant of 55.03° A—a deviation of only 9.97° from its objective slant, but 35.03° from the slant of the background.

In order to assess the relative magnitude of the slant-induction effect, the slant judgments were expressed as deviations from the slant of the background, and an analysis of variance was performed on these deviation scores. The analysis confirmed the observations recorded above. For any given degree of background slant the deviation was significantly greater for binocular than for monocular vision, i.e., the slant-induction effect was greater for monocular vision. The magnitudes of the deviations were not equal for all background slants within each condition of observation. The 20° T background slant produced the greatest deviation for monocular vision and the smallest deviation for binocular vision. However, if the large deviation for monocular vision may be assumed to be an exaggerated slant-induction effect, then we can conclude that the induction effect was more pronounced when both target and background were tilted forward.

Instructional sets.—An analysis of variance showed a significant interaction of attitude with condition of observation. When the target was viewed monocularly, instructional sets were ineffectual, i.e., they did not influence the shape-judgments. However, when the target was observed

TABLE 2

THE MEAN PROJECTIVE h/b RATIO REQUIRED BY THE APPARENT SLANT OF THE STANDARD COMPARED WITH THE MEAN PROJECTIVE h/b RATIO OF THE COMPARISON SETTING

Observation Attitude	Monocular			Binocular		
	Co. Mean Proj. h/b Ratio	St. Mean Proj. h/b Ratio	Sign and % Deviation	Co. Mean Proj. h/b Ratio	St. Mean Proj. h/b Ratio	Sign and % Deviation
	Background Slant = 20° A					
Phenomenal	.347	.200	+ 73.50	.507	.674	−24.77
Objective	.375	.251	+ 49.40	.656	.674	− 2.67
Analytic	.362	.343	+ 5.84	.541	.683	−20.79
	Background Slant = Vertical					
Phenomenal	.573	.609	− 5.74	.649	.735	−11.56
Objective	.556	.601	− 7.65	.710	.732	− 3.14
Analytic	.529	.608	− 12.17	.631	.733	−13.50
	Background Slant = 20° T					
Phenomenal	.368	.135	+172.59	.562	.409	+37.89
Objective	.367	.223	+ 64.12	.638	.514	+24.12
Analytic	.375	.208	+ 79.80	.483	.460	+ 5.00

binocularly, there were significant differences between the shape-judgments made under the various instructional sets. Here we will mention only that the objective attitude consistently resulted in matches which were significantly nearer to the objective dimensions of the target than the matches for the analytic and phenomenal attitudes. This finding is in agreement with the results reported by earlier investigators (Gottheil & Bitterman, 1951; Klimpfinger, 1933).

TABLE 3

SELECTED SAMPLE OF DATA FROM *S*s WHO DISPLAYED COMPLETE SLANT-INDUCTION WITH VERTICAL BACKGROUND AND MONOCULAR VISION

Observation Attitude	*N*	St. Objective h/b Ratio	Co. Mean Proj. h/b Ratio	St. Proj. h/b Ratio	Sign and % Deviation
Objective	10	.800	.540	.585	−7.68%
Analytic	9	.800	.543	.585	−7.17%
Phenomenal	7	.800	.555	.585	−5.12%
All attitudes	26	.800	.546	.585	−7.85%

Apparent slant-apparent shape invariance.—The following procedure was used to evaluate the invariance hypothesis. An h/b ratio was determined trigonometrically for each *S*'s *apparent* (judged) slant of the standard at each background slant. This ratio represents the projective ratio of the shape-match demanded for that degree of apparent slant by strict adherence to the invariance requirements. Next, the projective h/b ratio was determined for *S*'s setting of the comparison triangle at each background slant. A comparison of these two ratios allowed us to evaluate the degree of slant-shape invariance. Any difference between the two ratios represents a deviation from invariance requirements. The results of our analysis are summarized in Table 2.

To clarify the logic of this analysis a selected sample of data has been presented in Table 3. The mean h/b ratios contained in Table 3 were obtained from *S*s who manifested total slant induction when viewing the standard monocularly against a vertical background. In this special case, when the standard is perceived to be perpendicular to the line of sight, the frontal parallel projection for the apparent slant of the target is in effect the actual projective shape of the target. This is to say that, if the invariance requirements are to be strictly satisfied, the projective h/b ratio of the comparison (which in this special case is identical with the objective h/b ratio of the comparison) must equal the projective h/b ratio of the standard. For the dimensions of our standard this means that a perfect invariance fit would be h/b = 0.585. However, consulting the fourth column in Table 3, we find that this theoretical value was not attained. The deviations between the means were not great, but they were present, nonetheless.

An inspection of Table 2 reveals that in no instance did the mean projective h/b ratio of the comparison setting satisfy the invariance requirements. The deviations obtained for monocular observation ranged from −5.74% to 172.59%. For binocular vision the range of deviations was −2.67% to 37.89%. Relatively small deviations from invariance were obtained when the background was vertical. In addition, there was no significant difference between the grand mean deviations for monocular and binocular vision with the vertical background. The greatest deviations occurred when the background was slanted 20° T, although errors of considerable magnitude were obtained with the background 20° A. With only one exception, the deviations obtained with monocular observation under the 20° A and 20° T background conditions were at least twice as great as the deviations obtained for binocular vision.

A review of the ratio pairings for the individual *S*s (252 pairings, 126 for monocular and the same number for binocular vision) confirms these observations. There was not one case of perfect correspondence between the two h/b ratios. Table 4, which presents the distribution of deviations according to sign, reflects this fact.

Verbal reports.—The information derived from the interview tended to confirm the validity of the experimental operations. All *S*s gave evidence of having understood the task which was posed by the instruc-

TABLE 4

DISTRIBUTION OF DEVIATIONS FROM INVARIANCE REQUIREMENTS FOR MONOCULAR AND BINOCULAR VISION

Observation Attitude	Slant of Background						All Slants Deviations	
	20° A Deviations		Vertical Deviations		20° T Deviations			
	+	−	+	−	+	−	+	−
Monocular Vision								
Phenomenal	12	2	3	11	13	1	28	14
Objective	10	4	2	12	12	2	24	18
Analytic	8	6	3	11	13	1	24	18
All attitudes	30	12	8	34	38	4	76	50
Binocular Vision								
Phenomenal	3	11	3	11	11	3	17	25
Objective	6	8	8	6	11	3	25	17
Analytic	3	11	2	12	8	6	13	29
All attitudes	12	30	13	29	30	12	55	71

tions. Very few *S*s said that they had made deliberate efforts to take slant into account while making shape-judgments. All *S*s agreed that the judgments of slant made at the end of the experiment corresponded well with the apparent slant of the target during the earlier judgments of shape. Only a few *S*s expressed any strong lack of confidence in their judgments.

DISCUSSION

The results of this experiment confirmed and extended Beck and Gibson's (1955) findings concerning the slant-induction effect. The effect was observed to obtain not only when the background was vertical, but also when the background was slanted 20° A or 20° T. In fact, there is some reason for concluding that the effect was more pronounced when the background was slanted forward. We have no explanation for this latter finding, nor do we see any way of deriving it from the general stimulus-gradient theory.

The results with regard to the invariance hypothesis agreed with Beck and Gibson's findings in their general aspects, but were sufficiently different to warrant comment. The main point of disagreement was this: For both monocular and binocular observation our results showed much less adherence to the invariance requirements than did the results of Beck and Gibson. For instance, Beck and Gibson report that when the standard was viewed monocularly with a stationary head, all of their 30 *S*s selected a projective match, while with unrestricted binocular vision 23 of 30 *S*s selected an objective match. None of our *S*s made a perfect projective or objective match when these were demanded by the apparent slant.

The main reason for this disagreement resides in the difference between the measurement techniques employed in the two experiments.[4] We used a continuously variable comparison which enabled *S* to make sensitive discriminations. The method which Beck and Gibson used forced *S* to choose between one of two extreme alternatives, i.e., objective or projective. There was no way *S* could indicate any other perceived shape. In this situation *S* would probably select the comparison triangle which was *most like* the one he perceived even when he recognized that the two stimuli were not identical. Thus, all perceived shapes which clustered about the projective comparison were designated by that comparison and, similarly, all apparent shapes which clustered about the objective comparison were designated by that comparison. Lacking an opportunity to differentiate in the response system, *S*s produced results which appeared to reflect agreement in the perceptual system.

The evidence from past studies with regard to the effect of attitudes on the perception of shape is difficult to interpret. Klimpfinger (1933), who is most frequently cited, did not actually compare the influence of different attitudes within the same experimental situation. Instead, his conclusions were based mainly on a comparison of his data with the results of other *E*s. Gottheil and Bitterman (1951) used the same *S*s for each of the three attitudes. There is no way of knowing what effect judgments made under a previous attitude may have had on the later judgments obtained for another attitude. In addition, Gottheil and Bitterman do not tell us whether *S* viewed the target monocularly or binocularly. Our own results suggest an interaction of attitude with condition of observation. In light of the paucity of evidence on this question further research is needed.

SUMMARY

The experiment had three aspects: (*a*) to determine whether the slant-induction effect reported by Beck and Gibson (1955) would obtain when the background was slanted; (*b*) to test more precisely the slant-shape invariance hypothesis; and (*c*) to investigate the influence of three attitudes of

---

[4] Beck and Gibson (1955, p. 131) recognized the limitations of their technique for a precise test of the invariance hypothesis.

observation on the perception of shape under monocular and binocular vision.

The main results were: (*a*) The slant-inducing effect of the background was observed when the background was slanted 20° from the perpendicular either away from *S* (20° A) or toward *S* (20° T) in addition to the case in which the background was perpendicular to the line of sight. (*b*) Under no condition did the judgments of apparent slant and apparent shape covary exactly as demanded by the invariance hypothesis. The deviations from invariance ranged from −24.77% to 172.59%. (*c*) The influence of attitudes on the perception of shape was restricted to binocular viewing. Attitudes were found to be ineffectual in determining perceived shape when the standard was viewed monocularly.

REFERENCES

BECK, J., & GIBSON, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50**, 125–133.

GOTTHEIL, E., & BITTERMAN, M. E. The measurement of shape-constancy. *Amer. J. Psychol.*, 1951, **64**, 406–408.

KLIMPFINGER, S. Über den Einfluss von intentionaler Einstellung und Übung auf die Gestaltkonstanz. *Arch. ges. Psychol.*, 1933, **88**, 551–598.

KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.

LANGDON, F. J. The perception of changing shape. *Quart. J. exp. Psychol.*, 1951, **3**, 157–165.

LANGDON, F. J. Further studies in the perception of a changing shape. *Quart. J. exp. Psychol.*, 1953, **5**, 89–107.

NELSON, T. M., & BARTLEY, S. H. The perception of form in an unstructured field. *J. gen. Psychol.*, 1956, **54**, 57–63.

STAVRIANOS, B. K. The relation of shape-perception to explicit judgments of inclination. *Arch. Psychol., N. Y.*, 1945, No. 296.

(Received March 20, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 480–486

RETENTION OF CONNECTED DISCOURSE AS A FUNCTION OF DURATION OF INTERPOLATED LEARNING[1]

NORMAN J. SLAMECKA

University of Vermont

Several verbal learning studies have shown that recall of original learning (OL) varies inversely as a function of the degree of interpolated learning (IL) (Slamecka & Ceraso, 1960, p. 451). By *degree* of IL is meant the number of trials of presentation of a single IL list. Further, it is also known that recall of OL varies inversely with the amount of IL (Twining, 1940; Underwood, 1945). By *amount* of IL is mean the total number of different IL items presented. Amount of IL can be manipulated either by varying the number of different IL lists, or by varying the length of a single IL list.

The question whether amount of IL or degree of IL produces more forgetting has definite empirical value and theoretical interest, but to date there have been only two attempts to investigate it, and they both exhibit certain limitations. Underwood (1945) manipulated amount by varying the number of different IL lists. Using paired associates in an A-B, A-C paradigm, he compared three levels of degree of IL (8, 16, and 24 trials on a single list) with three levels of amount of IL (2, 4, and 6 different lists, with 4 trials on each list), in such a way that each level was equated for total trials. Results showed that retention decreased at a faster rate with increases in amount than with increases in degree of IL. However, it is rather ironic that in this unique experiment the usual progressive decrements in recall with increasing degrees of IL, which are commonly reported in other studies, were not obtained, and therefore the generalizability of the results is questionable. Had the expected progressive effectiveness of the degree variable been in fact obtained, the conclusions might have been quite different. As it is, the conclusion that with paired associates, amount produces more forgetting than degree must be regarded as tentative. McGeoch (1936), using serial lists of adjectives, reported the only study in which amount was manipulated by varying the length of the IL list. Interpolated lists were 8 or 16 units long, with presentation frequencies of 4, 8, and 16 trials. As in the Underwood study, differences in degree of IL within either of the list lengths failed to produce significant differences in recall. Further, although it was found that an 8-unit list given 8 or 16 trials produced slightly less retroactive inhibition (RI) than a 16-unit list given 4 or 8 trials, the differences were small and not statistically reliable. It is felt, therefore, that the question of the effects upon retention of these various ways of manipulating IL is still in need of investigation.

The present paper reports a series of three related RI experiments which seek to determine whether there is a common factor responsible for interference in the rote retention of certain serial materials, so designed as to allow direct cross comparisons among the experiments. Experiment I varied the degree of IL, Exp. II varied the amount by using different numbers of IL passages, and Exp. III used passages twice the length of those in the other two experiments. It was hypothesized that rote retention would be inversely related both to degree of repetition and to the number of IL passages given. Experiment III introduced a change in total trials across levels, since Exp. I and II were both comparable at each level in that regard. No hypotheses were formulated about any of the across-experi-

---

[1] This study was supported by National Science Foundation Grants G-6192 and G-14,721. The author is gratefully indebted to John A. Corson, who collected the data.

ment comparisons; however, the main intent and interest of the study is ultimately to focus upon those comparisons.

EXPERIMENT I

Method

Subjects.—Twenty-four undergraduate students of introductory psychology served as *S*s as part of the course requirements. Each *S* had five individual sessions, the first of which was a familiarization period designed to acquaint *S* with memory drum procedure and to lessen practice effects. This practice session involved the learning of a 20-word passage of connected discourse to one perfect trial, followed by six trials on an IL passage, and then two OL relearning trials. Each *S* was tested under all four experimental conditions.

Materials.—Serial materials consisting of 20-word passages of connected discourse were used. The following OL passages were presented:

1. Any generalized statement of relationship between variables requires specification of some reference class or universe from which the material comes.
2. According to recent considerations it appears preferable to select preliminary assumptions from the realm of objectively verifiable and observable phenomena.
3. Traditional rationalism cannot negate the assertion that its theory concerning primary evidence rests upon inadequate analysis of fundamental scientific methodology.
4. Forthright admission of the contingent character of our most soundly established beliefs distinguishes contemporary empiricists from dogmatic and positive ancients.

The following IL passages were used, of which each *S* was given three:

1. Early logical writings formulated the rationalist ideal of knowledge, but exploratory biological works exhibit less exacting standards of systematic inquiry.
2. Primitive conceptions of organization, such as classical geometry, were believed to be based upon completely demonstrable and absolutely indubitable principles.
3. Current convictions maintain that methods of natural science are the most reliable instruments thus far devised for comprehending empirical relations.
4. Next to establishment of the peripheral anchors of behavior, the observational tracing of mediating physiological processes has also merited investigation.
5. In the existing systems of measurement no contemporary laboratory investigator will state conclusions without stipulating their coefficient of probable error.
6. Psychology must insist upon greater precision than even the physicist may find necessary in his routine description of experimental situations.
7. The derivative modes of expression are more frequent in biology than in other natural disciplines, however, they definitely facilitate conceptualization.
8. The language of classical physics in the context of subatomic research is valuable in focusing attention directly upon important analogies.
9. Science rests upon assumptions, since every conclusion presupposes premises which are either primary or establishable by reliance upon other propositions.
10. These definitions contain all the necessary conceptual elements for making explicit the general characteristics of an objective approach to psychology.
11. Our knowledge is not infallible, but our general experience is evidence that even inconclusive formulations may differ in their adequacy.
12. Certain crucial expressions currently applied to quantum theory denote different traits of things than when used in discussing earlier formulations.

Design and procedure.—There were four degrees of IL: 0, 6, 12, and 18 trials, to be called, respectively, the Control, Low, Medium, and High. The standard RI design was used, consisting of OL (to a criterion of one perfect trial), IL, and relearning (RL) for 4 trials. The passages were presented on a memory drum at a 3-sec. presentation rate and a 6-sec. intertrial interval. OL passages and IL conditions were counterbalanced with respect to *S*s and sessions. IL passages were counterbalanced with respect to *S*s and conditions. The time from the end of OL to the start of RL was 21 min. The IL trials began 30 sec. after the end of OL, leaving, in the case of the Control, Low, and Medium conditions, periods of about 21, 14, and 7 min. free of any learning task. This time was filled by having *S* judge pictures from the Meier Art Judgment Test. To compensate for a loss in the set to respond to the memory drum, the above three groups also had one trial of number guessing for warmup just preceding RL.

Results

Mean OL trials to criterion in Exp. I (including the criterional trial) are found in Table 1. There were no significant differences among the means ($F = 1.14$, $df = 3/69$). Since interpolated trials were fixed, the IL acquisition is presented in terms of mean total correct anticipations, also shown in Table 1. It is evident that considerable differences were produced between the levels. Retention data for the first two RL trials in Exp. I are found in Table 2. By RL Trial 3 almost all *S*s were at a uniformly high level of relearning. Analysis of recall (RL Trial 1) showed a significant effect of degree of IL in the predicted direction ($F = 16.35$, $df = 3/69$, $P < .01$), and *t* tests between adjacent means showed significant differences between the Medium and Low IL conditions ($P < .05$, $df = 69$), and the Low IL and Control ($P < .01$, $df = 69$) conditions. These results support the experimental hypothesis and lend further corroboration to the findings of a previous study of degree of IL in connected discourse, wherein different acquisition criteria and different passages were used (Slamecka, 1960).

EXPERIMENT II

Method

Subjects.—Another 24 students of introductory psychology served as *S*s. As in Exp. I, each took part in five sessions, the first of which was a practice period. Each *S* had all four experimental conditions.[2]

Materials.—The same passages as those used in Exp. I were used in OL and in IL.

Design and procedure.—This experiment varied the number of different IL passages presented. The four levels used were zero, two, four, and six passages (to be called, respectively, Control, Low, Medium, and High), each being presented for three consecutive trials. The RI design was followed, with the same OL criteria, rate of presentation, intertrial interval, and OL-RL interval, warmup trial, filler task, and the four RL trials as in Exp. I. When the third trial on an IL passage was completed, *E* then exposed a new passage (after the normal intertrial interval) with no loss of tempo. The OL passages and IL conditions were counterbalanced with respect to *S*s and sessions. The IL passages were counterbalanced so that each one appeared an equal number of times within each condition.

Results

Mean OL trials to criterion for Exp. II are shown in Table 1. There were no significant differences among the means ($F = .66$, $df = 3/69$). Mean total correct anticipations during IL, also in Table 1, show obvious differences. Analysis of the recall means, found in Table 2, showed a significant effect of amount of IL ($F = 23.10$, $df = 3/69$, $P < .01$), and *t* tests indicated that all adjacent means differed significantly from each other in the predicted direction ($P < .01$ in all cases except Medium vs. High, where $P = .05$, $df = 69$). Therefore, the hypothesis that the retention of connected discourse is inversely related to the number of IL passages presented was given confirmation.

EXPERIMENT III

Method

Subjects.—Twenty-four students drawn from introductory psychology classes served as *S*s for five sessions, as previously described.

Materials.—The same OL and IL passages as in Exp. I and II were used. The only change was that the length of a given IL passage was doubled by linking two of the 20-word passages so as to form a two-sentence unit of 40 words to be learned. Thus, one trial with such a composite passage was equivalent, in elapsed time, to two trials with the shorter passages used previously. A total of six composite passages was obtained by combining the following numbered short IL passages: 1–7, 2–11, 3–4, 9–5, 10–6, 12–8.

---

[2] Two *S*s had to be replaced because of their inability to keep to the appointment schedule.

At each session, only one such passage was presented.

Design and procedure.—There were four levels of interference, designated by the number of trials given on a single composite passage. These were 0, 3, 6, and 9 trials, to be called, respectively, the Control, Low, Medium, and High conditions. It should be noted that the total presentation time for these passages, at each level, was equivalent to that in the other two experiments, since their increased length was compensated by halving the total trials. Essentially, this was a variant of the degree of IL study, and therefore the same experimental hypothesis would apply. The OL passages were counterbalanced with respect to sessions, *S*s, and conditions. The composite IL passages were counterbalanced with sessions, *S*s, and OL passages. Although complete counterbalancing of IL passages with conditions was not possible, a reasonable approximation to it was achieved. The general procedure followed was the same as in Exp. I and II.

Results

Mean OL trials for Exp. III are shown in Table 1. There were no significant differences among the means ($F = .99$, $df = 3/69$). Mean total correct anticipations during IL, in Table 1, show large differences. Mean amount recalled, as shown in Table 2, reflected significant differences attributable to degree of IL ($F = 16.64$, $df = 3/69$, $P < .01$), and t tests between adjacent means showed significant differences beyond the .01 level between the Low and Medium, and between the Medium and High conditions. Thus, degree of IL with the lengthened passages was also an effective variable. The significance of all three experiments when considered together will be discussed below.

TABLE 1

MEAN OL TRIALS TO CRITERION AND MEAN TOTAL CORRECT ANTICIPATIONS IN IL FOR ALL THREE EXPERIMENTS

Exp.	Control		Interpolated Learning					
			Low		Medium		High	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
	OL Trials							
I	10.6	3.4	9.5	3.1	10.8	3.0	11.0	3.7
II	10.4	5.2	10.9	5.2	10.2	3.8	11.5	5.4
III	10.1	3.8	11.3	3.6	10.5	4.7	11.5	5.0
	IL Anticipations							
I			70.3	9.5	188.4	15.2	304.3	21.4
II			40.0	12.4	77.2	21.8	121.7	29.5
III			29.1	9.7	96.7	26.8	195.0	42.6

TABLE 2

MEAN RETENTION SCORES FOR THE FIRST TWO RELEARNING TRIALS FOR ALL THREE EXPERIMENTS

Exp.	Control		Interpolated Learning					
			Low		Medium		High	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
	First RL Trial							
I	17.8	2.0	15.3	3.1	13.9	3.4	12.7	3.3
II	18.0	1.4	16.2	2.3	13.0	4.1	11.6	3.6
III	17.3	2.1	16.2	2.8	13.8	3.7	11.7	3.2
	Second RL Trial							
I	19.4	0.6	18.8	1.6	18.4	1.5	18.2	2.4
II	19.2	1.5	19.0	1.3	17.8	2.0	17.6	2.1
III	19.1	1.1	18.8	1.6	17.7	2.5	17.1	2.0

INTEREXPERIMENTAL COMPARISONS AND DISCUSSION

The reason for conducting three comparable experiments was to enable more precise information to be obtained regarding the specific source of the interference. Hitherto, verbal learning studies of the RI type have suffered from a form of confounding when seeking answers to this question. For example, the typical study of degree of IL employs varying numbers of trials on a single IL list. But, variation in the number of trials also results in concomitant variation in the total correct anticipations and in the total duration of the interfering learning as well. By the same token, a study of amount of IL in which the number of different lists is varied also involves concomitant variation in total IL time and total number of different IL items presented (see for example, Twining, 1940, where number of lists and number of items are completely confounded). This confounding can be circumvented by carrying out multiple, but nevertheless comparable experiments. The present study was designed so as to permit meaningful cross comparisons at each level, which would enable clearer identification of the factor or factors responsible for the forgetting.

At each of the three levels of interference there were wide differences in total trials, trials per passage, number of passages, number of different items, and mean total correct anticipations. Referring to Exp. I, II, and III in that order, we find that: (*a*) total trials were 6, 6, and 3 for Low; 12, 12, and 6 for Medium; and 18, 18, and 9 for High; (*b*) trials per passage were 6, 3, and 3 for Low; 12, 3, and 6 for Medium; and 18, 3, and 9 for High; (*c*) number of passages were 1, 2, and 1 for Low; 1, 4, and 1 for Medium; and 1, 6, and 1 for High; (*d*) number of different items were 20, 40, and 40 for Low; 20, 80, and 40 for Medium; and 20, 120, and 40 for High; and (*e*) mean total correct anticipations were 70, 40, and 29 for Low; 188, 77, and 97 for Medium; and 304, 122, and 195 for High. However, the one factor remaining constant within each level across all three experiments was the duration of IL, or the total time devoted to the interfering learning task. For the Control condition this was 0 min., for Low it was 6.6 min., for Medium it was 13.2 min., and for High, 19.8 min.

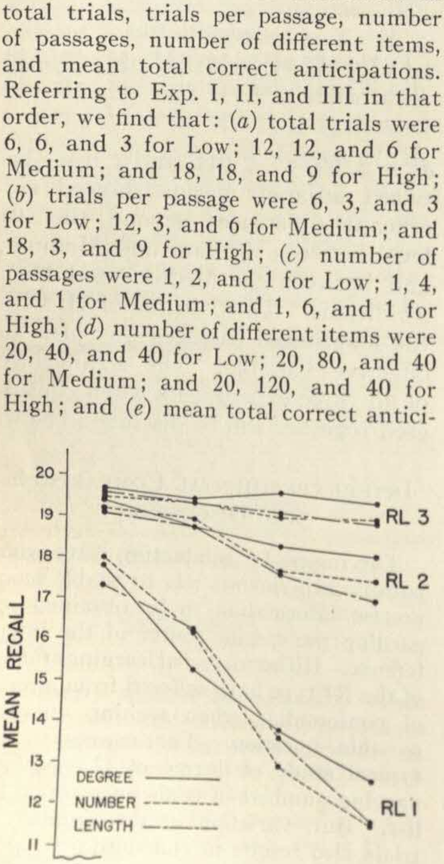


FIG. 1. Retention as a function of the time spent on the interfering learning task. (Control = 0 min., Low = 6.6 min., Medium = 13.2 min., and High = 19.8 min.)

The retention data have been plotted in Fig. 1 as a function of the IL durations. The Control conditions, where all *S*s had identical treatments, produced very similar recalls, testifying to the relative homogeneity of the *S*s used as well as to the degree of reliability of the techniques employed.

Analyses of the three recall (RL_1) means within each level gave *F*s of less than 1 in each case ($df = 2/69$), nor were there any within-level recall differences at the other relearning trials. Thus, equivalent amounts of forgetting were produced at each level by all three ways of manipulating the IL task. It made no difference how well the interfering items were learned, nor how many interfering items were presented, just as long as equivalent amounts of time were spent on the task. When forgetting was measured in terms of relative RI $\left[\frac{(\text{Rest-Work})}{\text{Rest}} \times 100\right]$, the percentages were: 14.0, 10.0, and 6.4 for Low; 21.9, 27.8, and 20.2 for Medium; and 28.6, 35.6, and 32.4 for High, indicating at least no consistent superiority of any one method over the others.

By virtue of the counterbalanced designs employed, the interference must be regarded as involving both RI and proactive inhibition (PI) together, since recent work has demonstrated that connected discourse is also subject to PI effects through the use of practiced *S*s (Slamecka, 1961). It might then be asked whether a similar pattern of retention scores would have been obtained had *S*s been naive. A partial approach to an answer with the present data would be to determine recall scores for performances on the first experimental day, thereby using *S*s with only the practice session as a possible source of PI. This results in a considerable loss of precision, since three-quarters of the original recall data would be unused. However, when calculated in this fashion, the recall means for Exp. I, II, and III, in that order, were: 19.0, 17.8, and 18.7 for Control; 14.3, 16.0, and 16.3 for Low; 15.3, 12.2, and 15.2 for Medium; and 13.5, 13.5, and 13.7 for High. Thus, even with only 6 *S*s per data point, there is no evidence of a consistent difference in recall for any one experiment over the others, and under conditions of maximal interference the recalls are practically identical. Our main results, therefore, do not appear to be specific only to highly practiced or PI-burdened *S*s.

For purposes of analysis, errors at recall were classed either as omissions, intralist intrusions, or "other" intrusions. In view of the small number of errors comprising the other class, it did not seem justified to attempt a further subclassification of them. Omissions accounted for three-quarters of all errors, and within-list sources accounted for two-thirds of the intrusion errors. Both classes of intrusions tended to rise with increasing interference and to maintain a relatively stable ratio with respect to each other in all three of the work conditions. Furthermore, there was no apparent trend in these ratios across the experiments. The ratios of within-list to other intrusions, in terms of mean proportions across experiments, were: 3.2 for Control, 1.7 for Low, 1.9 for Medium, and 1.8 for High. The ratios of within-list to other intrusions in terms of means across conditions were: 1.7 for Exp. I, 2.1 for Exp. II, and 2.1 for Exp. III. It appears that intrusion analysis does not reveal any characteristic differences among the three experiments. The percentages of all overt responses made that were correct decreased from the Control condition to the High IL condition. The average percentages over the three experiments were: 94.7 for Control, 93.7 for Low, 91.2 for Medium, and 87.2 for High. There was no significant variation in these percentages between experiments.

It is likely that with connected discourse, even the errors committed would be shaped by the contextual constraints and the sequential dependencies in the language. Many of the other intrusions were synonyms of the correct word, or else were otherwise appropriate to the context of the given passage. The within-list intrusions appeared to be largely anticipatory, as if *S* had skipped over some of the words, and such errors were also usually contextually appropriate.

The above analyses of recall, relearning, errors, and accuracy percentages, imply a quite clearcut principle, namely, that the rote retention of passages of connected discourse subjected to interpolated interference is solely a function of the amount of time spent on the interfering task, regardless of how that task is structured. There is no a priori reason why these results should not be as readily obtainable with serial lists of an unconnected or nonsense type, and if comparable findings are reported with such materials we would have a useful simplifying generalization about rote serial retention.

This study also has some bearing on explanations of forgetting which stress confusion of list membership, according to the notion that the more interfering material that is presented, the greater the opportunity for confusion among items from different lists. In Underwood's (1945) study, where the A-B, A-C list construction is designed for specific interference, such an hypothesis appears quite reasonable. But serial materials may not involve specific item interference, and thus the amount of interfering material would not necessarily result in changes in confusion of list membership. In the present study, the conditions of Exp. II should have maximized such confusion, since many IL passages were presented, each to such a low degree that differentiation would be quite weak. Nevertheless, retention was no less in Exp. II than in Exp. I, where only a single IL passage was given, and for many more trials. Rather, it seems that the interference was nonspecific with regard to the form of the IL task, its magnitude determined by the time spent on that task, and the overt errors governed largely by local contextual constraints.

SUMMARY

Three related RI experiments were made to determine more precisely the source of interference in the rote retention of passages of connected discourse. Experiment I varied the degree of IL by giving 0, 6, 12, or 18 trials on a single passage; Exp. II varied the number of different IL passages by presenting 0, 2, 4, or 6 passages, each for three consecutive trials; Exp. III used passages twice as long as those given in the other two experiments, for half the number of trials, viz., 0, 3, 6, and 9. It was hypothesized that retention would be inversely related both to the degree of learning and the number of IL passages presented. The main intent of the study, however, was an across-experiment comparison at each of the levels. The one common factor within each level was the IL duration which was, for each interference level, 0 min., 6.6 min., 13.2 min., and 19.8 min. The same passages were used in all experiments.

Results were as follows: (*a*) Exp. I and III verified the hypothesis that retention was inversely related to degree of IL; (*b*) Exp. II confirmed the hypothesis that retention was inversely related to amount of IL; (*c*) Cross-experiment comparisons at each level showed no significant differences at recall or relearning; (*d*) the pattern of errors at recall was rather comparable for all experiments, with most errors being omissions, and the bulk of intrusions coming from within the passage. Contextual constraints seemed to determine the kinds of overt errors made.

It was concluded that the rote retention of passages of connected discourse subjected to interpolated interference is solely a function of the amount of time spent on the interfering task, regardless of how that task is structured. The data failed to support notions stressing confusions of list membership at recall, but rather pointed to a nonspecific interference effect for serial materials.

REFERENCES

McGeoch, J. A. Studies in retroactive inhibition. VII. Retroactive inhibition as a function of the length and frequency of presentation of the interpolated lists. *J. exp. Psychol.*, 1936, **19**, 674–693.

Slamecka, N. J. Retroactive inhibition of connected discourse as a function of practice level. *J. exp. Psychol.*, 1960, **59**, 104–108.

Slamecka, N. J. Proactive inhibition of connected discourse. *J. exp. Psychol.*, 1961, **62**, 295–301.

Slamecka, N. J., & Ceraso, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

Twining, P. E. The relative importance of intervening activity and lapse of time in the production of forgetting. *J. exp. Psychol.*, 1940, **26**, 483–501.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59** (3, Whole No. 273).

(Received March 20, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 487–494

EFFECTS OF INSTRUCTIONAL SET AND UCS INTENSITY ON THE LATENCY, PERCENTAGE, AND FORM OF THE EYELID RESPONSE[1]

I. GORMEZANO AND JOHN W. MOORE

Indiana University

The purpose of the present investigation was to compare the effects of voluntary and nonvoluntary blink instructions on the latency, frequency, and form of eyelid responses under classical conditioning and extinction procedures. The study was also concerned with determining the effects on these response variables of independently varying UCS intensity.

The procedure of the Iowa eyelid conditioning laboratory of rejecting *S*s on the basis of response latency assumes there is a causal relationship between responses characterized by sharp closure, long duration, and short latency, and the voluntary attempt of *S* to blink to the CS (Spence, 1953; Spence & Ross, 1957, 1959; Spence & Taylor, 1951). Spence and Ross (1959) have attempted to provide analogous support for this assumption by maintaining that judged voluntary responses are ". . . similar in form to those given by *S*s instructed to blink to the CS so as to avoid the air puff and by *S*s who, following conditioning, reported they were blinking voluntarily in order to avoid the puff" (p. 377). (Presumably, *S*s "avoiding" the air puff are attempting to attenuate the noxious effects of the UCS by having it fall on the closed eyelid.) Explicitly, Spence and Ross (1959) have stated ". . . the rationale behind this procedure of eliminating *S*s whose responses were primarily of the short latency, voluntary form was that such responses were governed by different laws than those obtaining in the case of the longer latency class of responses" (p. 377). As yet, Hartman, Grant, and Ross (1960) have reported the only study which has attempted to examine the response characteristics of *S*s instructed to blink to the CS to "avoid" the air puff. These investigators examined the latency distributions of *S*s instructed to blink voluntarily under ready versus no ready signal and massed versus spaced trial conditions and reported qualitative correspondence with the latency distributions of judged voluntary *S*s recovered from the data of Spence and Ross (1959) and Hartman and Ross (1961).

Since the latency distribution of judged voluntary *S*s is itself recovered from eyelid conditioning data, the possibility exists that the latency distribution, and other response characteristics of judged voluntary *S*s, would show some degree of correspondence with conditioning data obtained under a wide variety of independent variables. In the present investigation orthogonal comparisons were made of the response characteristics of voluntary blink instructed *S*s with that of nonvoluntary blink instructed *S*s (i.e., *S*s given conventional "neutral" instructions). Further, the present investigation attempted to determine if variations in UCS intensity had any differential effects on the response characteristics of *S*s under the two instructional sets.

---

[1] This research was supported in part by a grant from the National Institute of Mental Health, Grant M-4034(A) and Grant G 16030 from the National Science Foundation. A brief report of the experiment was presented at the Psychonomic Society, Chicago, September 1960.

METHOD

Apparatus.—The apparatus was the same as that described in detail by Moore and Gormezano (1961). The *S*'s room was sound treated and contained two identical enclosures which permitted 1 or 2 *S*s to be run in a session with identical stimulating and recording components. The *S* sat with his head comfortably positioned on a chin rest and facing into the enclosure. About 3 ft. in front of his eyes was a 10-cm. circular milk glass window. An abrupt change in the illumination of this window, from .19 to .26 mL., served as the CS. Also present in the enclosure was an intercommunication set and two 15-w. incandescent lamps. The *S* wore an elastic headband supporting an air-jet with a 2-mm. diameter orifice and a microtorque potentiometer used to pick up movement from his right eyelid. The signal from the potentiometer was amplified by a Brush dc amplifier and recorded by means of an ink-writing Brush penmotor. Brush penmotors also recorded the onset and termination of the CS and UCS. A Western Union tape transmitter controlled the intertrial intervals and activated electronic timers controlling the durations of the CS and UCS, and the CS-UCS interval. The duration of the CS was 600 msec. and the CS-UCS interval 500 msec. The UCS was a puff of compressed nitrogen delivered to the right cornea and terminated with the offset of the CS. The intensities of the air puffs, measured at their point of delivery to the eye, were sufficient to support a 50-, 100-, or 200-mm. column of mercury.

Design.—Twenty-four *S*s were randomly assigned to each of the six cells of a 2 × 3 factorial design in which voluntary (V) and nonvoluntary (NV) blink instructions were made orthogonal to three UCS intensities of 50, 100, and 200 mm. The only restriction upon the random assignment of *S*s was that the proportion of women to men in each of the six cells remain the same (8 women: 16 men). The two recording systems were also made orthogonal to the instruction and UCS intensity dimensions so as to remove their contributions to the error variance.

Procedure.—Instructions appropriate to *S*'s group were given over the intercommunication set. The instructions given to the nonvoluntary blink groups (NV50, NV100, and NV200) were as follows:

> Please listen carefully to the following instructions. Remain seated comfortably and keep looking at the milk-white disk in front of you. Do not touch anything on your head at anytime during the experiment.
>
> You will see and feel a series of stimuli during the experimental session. The stimuli will consist of a light and a puff of air. Be careful not to control voluntarily your natural reactions to the stimuli. Do not try to figure out the experiment. Keep as detached an attitude as possible and simply let your reactions take care of themselves.
>
> You can communicate with me at anytime by speaking in a normal voice. Are these instructions perfectly clear to you? Please tell me what you are to do.

The only difference in the instructions given to *S*s in the voluntary blink instructed groups (V50, V100, and V200) occurred in the second paragraph which was given as follows:

> You will see and feel a series of stimuli during the experimental session. The stimuli will consist of a light and a puff of air. Your task is to blink so as to "avoid" the puff of air.

Following the instructions all *S*s received 70 acquisition and 20 extinction trials. The intertrial interval for the 90 trials was randomized at 15, 25, and 35 sec. with a mean of 25 sec. The response characteristics of eyelid closures of at least 1-mm. deflection from the baseline, in the latency range of 150 to 525 msec. during acquisition and 150 to 600 msec. during extinction, were those subjected to statistical analyses.

Subjects.—The *S*s were 96 men and 48 women from introductory psychology courses at Indiana University. In addition, 6 men and 10 women were discarded from the experiment because of apparatus failure or experimental error.

RESULTS

The frequency distribution of response latencies in acquisition is shown in Fig. 1 for all six groups. The V groups are depicted on the left and the NV groups on the right with UCS intensity as the parameter. Each of the distributions were clearly bimodal. The responses in the first modes from the left were distributed about a 75-msec. mode and are commonly taken as reflex responses to a visual CS (Grant, 1943). The low frequency of responses in the 125- to 200-msec. interval among the NV groups would seem to indicate that precautions to maintain light adaptation were successful in eliminating Beta responses (Grant & Norris, 1946). Concerning the principal modes appearing on the right, it is clear that instructional set had a considerable effect on response latency. The modal response occurred before 300 msec. in the V groups and after 300 msec. in the NV groups. Further examination of the distributions indicates that as UCS intensity increased the modal response latency decreased under both instructional sets.

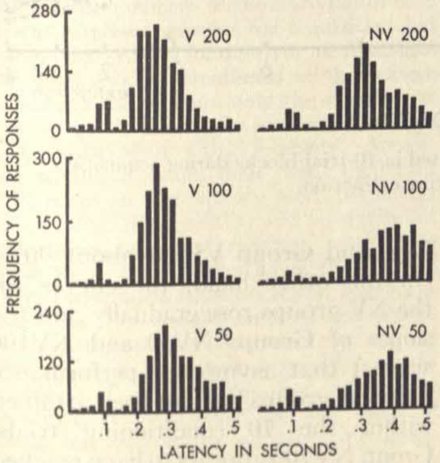


FIG. 1. Latency distribution of all eyelid responses in acquisition.

Figure 2 shows the frequency distribution of response latencies in extinction. Though all groups showed a drop in frequency of responses, the striking aspect of these distributions is that the modal responses of the V groups did not appear to have changed to any considerable extent from acquisition to extinction. On the other hand, the NV groups demonstrated particularly sharp decreases in response frequency around the modal response latency observed in acquisition (i.e., the 300- to 400-msec. interval).

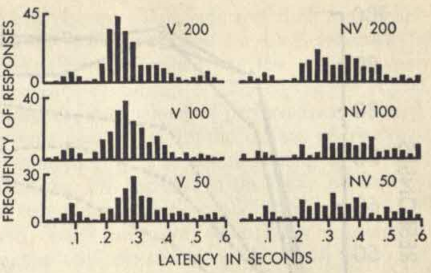


FIG. 2. Latency distribution of all eyelid responses in extinction.

Table 1 presents the effects of instructional set and UCS intensity on the latency and form of eyelid responses during acquisition and extinction. The cell entries for latency are the mean latency of responses in milliseconds and those for angle are the mean response angles in degrees. The response angle was determined from the graphic records by taking the angle formed between the baseline and the line passing through the points of response onset and maximum lid closure in the bounded interval of 150 to 525 msec. in acquisition

TABLE 1

MEAN LATENCY AND ANGLE OF EYELID RESPONSES DURING ACQUISITION AND EXTINCTION

Group	Latency (Msec.)		Angle (Deg.)	
	Acquisition	Extinction	Acquisition	Extinction
V50	331.2	331.6	28.3	25.9
V100	297.9	287.5	27.6	23.3
V200	284.8	288.6	27.8	23.8
NV50	384.3	399.6	20.3	21.9
NV100	390.7	400.0	17.7	17.2
NV200	351.6	358.0	19.9	20.6

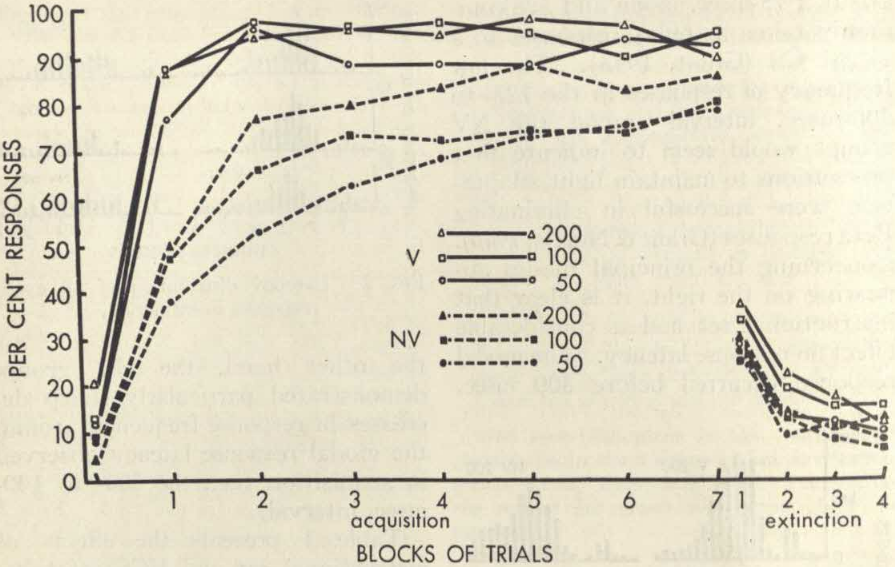


FIG. 3. The percentage of responses plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction.

and 150 to 600 msec. in extinction. This measure represents an attempt to objectively quantify the form of eyelid closures. From examination of the table it is apparent that in acquisition and extinction, the V groups had considerably shorter mean latencies and sharper lid closures than the NV groups. It is also clear that as UCS intensity increased from 50 to 200 mm. the mean latency of responses decreased under both instructional sets, and response angle failed to show systematic change.

Figure 3 presents the performance curves in acquisition and extinction. The initial points for all acquisition curves are the mean percentage of responses on Trial 1 and the remaining acquisition points are the mean percentage of responses for each 10-trial block. The extinction curves are plotted in 5-trial blocks. In acquisition, the V groups proceeded to asymptote within 20 trials. Groups V100 and V200 stabilized at around 95% and Group V50 at about 90%. On the other hand, the curves for the NV groups rose gradually and the slopes of Groups NV50 and NV100 suggest that asymptotic performance for these groups had not been attained within the 70 conditioning trials. Group NV200 appears to have reached an asymptotic level of about 85%. The mean percentage of responses for Groups NV50, NV100, and NV200 in acquisition were 64.1, 69.0, and 78.0%, respectively, whereas the means were 88.4, 93.9, and 93.2% for Groups V50, V100, and V200, respectively. For the extinction curves presented on the right hand side of the figure, it is of particular interest to note that the higher level of responding of the V groups in acquisition also persisted in extinction. The mean percentage of responses in extinction for Groups NV50, NV100, and NV200 were 28.8, 27.1, and 34.2%, respectively, and for Groups V50, V100, and V200 the means were 32.5, 40.4, and 45.4%, respectively. Thus, under each instructional set there was a small but increasing resistance to extinction as UCS intensity increased.

Statistical analysis of each of the response measures was carried out by analysis of variance. The latency and form analyses were performed on the log transforms of individual mean response latencies and individual mean response angles, respectively; the analysis of percentage of responses was performed on the arc-sine transform of individual percentage scores. These transformations served to bring about homogeneity of variance, making them more amenable to the analysis of variance (Bartlett, 1947). Table 2 gives the summary of the analysis for each of the response measures for acquisition and extinction. Except for the error mean squares which appear in parentheses at the bottom of each of the columns, only the *F* ratios are presented. In Row 1, the source of variation tested is the effect of UCS intensity. According to the *F* values there were significant differences in the latency and percentage of responses in acquisition and no significant differences for either response measure in extinction. On the other hand, the analyses of response angle failed to reveal significant differences in acquisition and extinction. Individual *t* test comparisons of the arc-sine transform of percentage acquisition scores yielded a *P* value of less than .01 for the NV50 v. NV200 comparison and failed to reach the .05 level for all other intergroup comparisons. This indicates that the significant *F* value obtained for UCS intensity is primarily determined by the superior performance of Group NV200 to Group NV50. Since the mean level of performance of the V groups was close to the ceiling of response measurement, it is not surprising to observe that the 5% increase in the mean percentage of responses from Group V50 to Groups V100 and V200 was not statistically significant. In the analyses of the log transform of latency acquisition scores, *t* test comparisons yielded *P*s of less than .05 for four of the six intergroup comparisons (viz., V50 vs. V100, $P < .02$; V50 vs. V200, $P < .001$; NV50 vs. NV200, $P < .05$; NV100 vs. NV200, $P < .01$). The failure of the NV50 vs. NV100 *t* test comparison to reach statistical significance indicates that the observed increase in response latency from Group NV50 to Group NV100 was not a significant inversion in the observed UCS intensity-response latency function. Row 2 indicates that voluntary blink instructed *S*s had significantly shorter mean response latencies, higher levels of percentage of responses, and sharper lid closures than nonvoluntary blink instructed *S*s in both acquisition and extinction. In Row 3 it is shown that the source of variation attributable to differences in the two recording systems failed to reach significance for any of the response measures in both acquisition and extinction. Row 4 presents the interaction effects of UCS intensity and instructional set. The failure of the *F* values to reach statistical significance in acquisition or extinction suggests that variations in UCS intensity had

TABLE 2

ANALYSES OF VARIANCE OF LATENCY, PERCENTAGE, AND ANGLE OF EYELID RESPONSES DURING ACQUISITION AND EXTINCTION

Source of Variation	*df*	Latency		Percentage		Angle	
		Acquisition	Extinction	Acquisition	Extinction	Acquisition	Extinction
UCS intensity (I)	2	9.85***	2.93	4.29*	1.26	.47	1.48
Instructional set (S)	1	94.32***	58.26***	82.98***	4.94*	26.12***	6.85*
Recording system (R)	1	.06	.11	.53	.10	1.32	1.97
I × S	2	2.77	2.85	1.55	.45	.70	.77
I × R	2	.65	.82	.54	1.19	.48	.46
S × R	1	1.94	.61	.07	.08	6.67*	4.92*
I × S × R	2	1.70	3.70*	.06	1.94	.16	.17
Error	132	(.0034)	(.0061)	(112.17)	(443.52)	(.0405)	(.0598)

Note.—Number in parentheses are error mean squares. The latency and angle analyses were performed on the log transforms of individual mean latencies and individual mean response angles; the frequency analyses were performed on the arc-sine transforms of individual percentage scores.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

similar effects under both instructional sets. Examination of the remaining *F* values in the table indicates that significant *F*s were obtained in Rows 6 and 7. Row 6 shows there were significant interaction effects of instructional set and recording system on the form of the response in acquisition and extinction. These interaction effects arose from the fact that in both acquisition and extinction the mean response angle under V instructions was greater in Recording System I than in Recording System II, whereas under NV instructions the mean response angle was greater under Recording System II than under Recording System I. The significant *F* value in Row 7 indicates that the second-order interaction of UCS intensity, instructional set, and recording system had a significant effect on response latency in extinction. However, the significant *F* values obtained in Rows 6 and 7 appear to be devoid of psychological significance.

DISCUSSION

The principal findings of this investigation were: (*a*) *S*s under voluntary instructional set exhibited responses of shorter latency, attained higher levels of percentage of responses, and had sharper lid closures than *S*s under nonvoluntary instructional set in both acquisition and extinction; (*b*) as UCS intensity increased it was observed that in acquisition response latency decreased, percentage of responses increased, and sharpness of lid closure showed no systematic change; and (*c*) there were no significant interaction effects between instructional set and UCS intensity on the latency, percentage, or form of the response in acquisition or extinction.

The first finding indicates that the response characteristics of voluntary blink instructed *S*s are markedly different from those of *S*s conditioned under conventional neutral instructions. It also indicates, so far as acquisition is concerned, considerable similarities between the response characteristics of voluntary blink instructed *S*s and those attributed to judged voluntary responders by Spence and his associates (Spence, 1953; Spence & Ross, 1957, 1959; Spence & Taylor, 1951). The latency distributions of voluntary blink instructed *S*s further confirms the correspondence to judged voluntary responders reported by Hartman, Grant, and Ross (1960) and are also strikingly similar to judged voluntary responses without a ready signal, reported by Hartman and Ross (1961). The observation that the response characteristics of voluntary blink instructed *S*s were maintained in extinction suggests that their responses became conditioned under the acquisition procedure. (An alternative possibility is that voluntary blink instructed *S*s interpreted the instructions to require that they blink to the CS even in the absence of the UCS.) As yet, conditioning studies which have recovered judged voluntary responders from their data have not attempted to investigate the behavior of these *S*s in extinction. If however, their extinction curves showed a greater resistance to extinction, it would suggest that the term "voluntary responder" is very inappropriate and that "highly conditionable" *S*s (i.e., *S*s with higher learning rate parameters) would be a better one.[2]

The second finding, that increasing UCS intensity results in increasing percentage of responses in acquisition, is a conventional result. However, the fact that increasing UCS intensity produces systematic decreases in response latency has methodological implications for eyelid conditioning studies which have varied UCS intensity and concurrently employed a latency criterion to reject *S*s as voluntary responders. As UCS intensity increases such a procedure would classify increasing numbers of *S*s as voluntary responders. (The possibility also exists that the observed latency-UCS intensity relationship would be different for studies employing a ready signal.) It could be maintained, of course, that increases in UCS intensity simply increases the incidence of voluntary responders. Spence and Ross (1959) implied this argument when they stated that ". . . a relatively strong air puff was employed in order to elicit a large number of voluntary type responses" (p. 378). Such reasoning brings into sharp focus another difficulty in current attempts to classify *S*s as voluntary responders on the basis of the topography of their responses. Unless it has been empirically determined that the response characteristics attributed to voluntary responders, and employed as criteria in their selection, are *not* a function of independent variables employed in the conditioning procedure, the equally valid assumption can be made that *S*s will be rejected as voluntary responders for behavior lawfully related to different independent variables and/or different values of the independent variables. Spence and his associates may choose to delete these *S*s on the basis of response latency in order to simplify their descriptive problem within the Hull-Spence theoretical framework. However, the basic difficulty is that one is forced into the logically difficult position of having to employ criteria in the selection of these *S*s which have been shown to be invariant, not only for a restricted set of conditioning parameters, but for the infinite population of parameters. In the absence of such cross validation of these criteria, one is confronted with the possibility of erroneously classifying (or failing to classify) different numbers of *S*s as voluntary responders for behavior functionally related to the conditioning parameters employed. The failure of response latency to satisfy this condition of invariance is indicated by the presently observed latency-UCS intensity relationship and further, by the findings of Hartman, Grant, and Ross (1960) that the response latency of voluntary blink instructed *S*s remained invariant for massed and distributed conditioning trials with a ready signal, but that a shorter response latency was observed in massed than spaced conditioning trials in the absence of a ready signal. Whether objective measures of sharpness of lid closure will meet with similar difficulties remains an empirical question.

---

[2] To explore such a possibility, the latency criterion suggested by Spence and Ross (1959) was applied to *S*s in the NV groups (i.e., *S*s who gave 50% or more of their CRs before 300 msec. were classified as "voluntary" responders). Since a ready signal was not employed in the present investigation this latency criterion would presumably be a conservative criterion for recovering voluntary responders (Hartman et al., 1960). Nevertheless, 3 voluntary *S*s were recovered from Group NV50, 2 from NV100, and 3 from NV200. In acquisition, the mean percentage of CRs for these *S*s were 84.8% for NV50, 86.4% for NV100, 90.8% for NV200, and in extinction the means were 66.7, 47.5, and 49.0%, respectively. Though these observations were based on a small and unequal number of *S*s, the mean percentage of CRs of these voluntary *S*s was considerably greater in both acquisition and extinction than that obtained under each of the NV conditions with unselected *S*s.

The third finding of failing to obtain a significant interaction effect of instructional set and UCS intensity on percentage of responses is related to the observation of Spence and Ross (1959) that judged voluntary responders follow different laws of conditioning than nonvoluntary responders. Specifically, it has been maintained that the percentage of responses of these *S*s do not vary with changes in UCS intensity. However, if judged voluntary *S*s are simply *S*s with higher conditioning parameters and they are selected from experiments in which there were differences in UCS intensity, then one might expect to observe no differences since they are performing near the ceiling of response measurement. The fact that the response angles of voluntary blink instructed *S*s followed similar laws for variation in UCS intensity as nonvoluntary instructed *S*s suggests that a specified criterion of response angle could be employed to discriminate the two instructional groups independent of UCS intensity. The observed invariance requires proof of the null hypothesis. Although this cannot be done, it is proper to indicate that with 144 *S*s the power of the experiment was considerable, so that if the latency, frequency, and form of responses under the two instructional sets did follow different laws for UCS intensity the differences must not have been appreciable.

SUMMARY

The effects of voluntary (V) and nonvoluntary (NV) blink instructions on the latency, frequency, and form of eyelid responses were compared under classical conditioning and extinction procedures, with UCS intensity as a parameter. It was found that: (*a*) The *S*s under V instructions gave responses of shorter latency, attained higher percentages of responses, and had sharper lid closures than *S*s under NV instructions in both acquisition and extinction. (*b*) As UCS intensity increased, acquisition response latency decreased, percentage of responses increased, and the form of lid closure failed to show systematic change. (*c*) There were no significant interaction effects between instructional set and UCS intensity.

It was concluded that: (*a*) the responses of the V *S*s became conditioned under the acquisition procedure; (*b*) studies that have varied UCS intensity and concurrently employed a latency criterion to reject *S*s as voluntary responders may have introduced a methodological error; (*c*) V *S*s may not follow appreciably different laws to variations in UCS intensity than NV *S*s.

REFERENCES

BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.

GRANT, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32**, 201–212.

GRANT, D. A., & NORRIS, E. B. Dark adaptation as a factor in the sensitization of the Beta response of the eyelid response to light. *J. exp. Psychol.*, 1946, **36**, 390–397.

HARTMAN, T. F., GRANT, D. A., & ROSS, L. E. An investigation of the latency of "instructed voluntary" eyelid responses. *Psychol. Rep.*, 1960, **7**, 305–311.

HARTMAN, T. F., & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 334–338.

MOORE, J. W., & GORMEZANO, I. Yoked comparisons of instrumental and classical eyelid conditioning. *J. exp. Psychol.*, 1961, **62**, 552–559.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

SPENCE, K. W., & ROSS, L. E. Experimental evidence on the relation between performance level in eyelid conditioning and anxiety (drive) level. *Off. Naval Res. tech. Rep.*, 1957, No. 5. (Contract No. N9-ONR-93802).

SPENCE K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

SPENCE, K. W., & TAYLOR, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

(Received March 22, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 495–498

TENDENCIES TO EYE MOVEMENT AND PERCEPTUAL ACCURACY[1]

HERBERT F. CROVITZ AND WALTER DAVES

Veterans Administration Hospital, Durham, N. C.

Duke University

The present study deals with one possible role of sensory-motor factors in perception. While the movement of the eyes leads to new visual stimulation which may be relevant in perceiving, the neural activity responsible for an eye movement may itself play a role in perceptual integration whether or not an eye movement occurs and whether or not the sensory feedback effects of an eye movement are supplied to the locus of perceptual integration. This neural state tending to produce an eye movement and having a possible role in perceptual processing will be called a "tendency to eye movement."

The postulated process may be related to the process underlying the use of the term tendencies to eye movement as used previously in psychology. For example, Woodworth (1938) cites a theory relating eye movements to the Muller-Lyer illusion which ". . . admits that actual movements do not occur in all cases, but assumes that a tendency to such movement is sufficient to give the impression of length" (p. 645). Recently the term has been used in perceptual theory by Gaffron (1950), Heron (1957), and Crovitz (1960).

In normal visual experience, perception may be affected by both the feedback supplied by eye movement and by the neural state which initiated the movement. Using a tachistoscope, however, when exposure time is shorter than the latency of eye movement, the eye movements which occur lead to the fixation of the blank postexposure field. The *first* eye movement which occurs must be the result of the tendency to eye movement existing at the initiation of the movement. If, as assumed, there is a relationship between the tendency to eye movement qua neural state, and perception, a relationship should exist between some property of the initial eye movement, as an index of the tendency, and some property of perception. It would appear reasonable that a property in which a relationship could be shown would be the directionality of eye movement and perceptual accuracy; i.e., a congruence might be expected between the direction of the initial eye movement and the direction of the more accurate field.

METHOD

Stimulus materials.—The stimulus material[2] consisted of three sets of eight cards, each card with numerals appearing at 3°, 5°, and 7° of visual angle of both the left and right stimulus fields of a Gerbrands mirror tachistoscope. The tachistoscope was fitted with an

---

[1] The authors are indebted to Karl Zener for his valued advice in the formulation and execution of this study, to Robert L. Green who made available part of the eye movement recording apparatus used, and to Paul G. Daston for critical reading of the report.

[2] The perceptual situation chosen was determined by requirements of a replication of an unpublished study on perceptual anisotropies and laterality done by Hilborn at Duke University. The present report, however, does not relate to the specifics of Hilborn's problem.

TABLE 1

FREQUENCY OF OCCURRENCE OF EACH COMBINATION OF ACCURACY DIFFERENCE AND EYE MOVEMENT DIRECTION FOR EACH VIEWING CONDITION

Viewing Condition	More Accurate Field	Eye Movement Direction			
		Rt.	Lt.	No.	Total
Both eyes	Right	30	3	10	43
	Left	2	32	7	41
	Equal	5	14	9	28
	Total	37	49	26	
Right eye	Right	39	4	12	55
	Left	6	8	13	27
	Equal	8	11	11	30
	Total	53	23	36	
Left eye	Right	18	8	11	37
	Left	4	26	13	43
	Equal	3	17	12	32
	Total	25	51	36	

Note.—Rt. = Right; Lt. = Left; No. = None.

eye aperture occluder such that the line of numerals across the whole field could be viewed binocularly or with the left or the right eye alone. Each set of eight cards used in the three viewing conditions (both eyes, left eye alone, right eye alone) had each of the numerals (2–9) in each of the six positions once. Otherwise the positions for the numerals were randomly selected. The numerals, executed in India ink using a Zephyr lettering kit, stood about $\frac{3}{16}$ in. high, and subtended a visual angle of about 30′.

Procedure.—Cards were presented in a fixed random order, binocular and monocular trials intermixed. On each trial *S* adjusted the eye aperture occluder according to *E*'s instructions, viewing with both eyes, left eye alone, and right eye alone for a total of eight trials each. A constant exposure time of 100 msec. was used, and since the latency of voluntary eye movements is somewhat longer than this (Woodworth, 1938) such post-exposure eye movements could not lead to fixation of a numeral.

Recording technique.—An electro-oculograph technique was used to record initial postexposure eye movements. Eye movement potentials were amplified by a Grass Model III-D high efficiency EEG machine equipped with converter-demodulators (Grass Model CD-3) in order to provide dc voltage input to the pen. Electrodes were made of jeweler's "high fine" silver disks, 1 mm. thick by 8 mm. diameter, plated with silver chloride and encased in silver cups, in general agreement with the method suggested by Ford and Leonard (1958).

Electrodes were placed bitemporally for recording the horizontal components of the eye movements. A second channel was connected to the tachistoscope timer such that a sharp spike was produced on the moving paper record each time the stimulus field was flashed on and each time it was flashed off, allowing for and leading to a check that the initial eye movement recorded began after the stimulus material flashed off. Paper speed was 30 mm/sec.

Before the electrodes were taped into position, the skin sites were washed with acetone and lightly sanded with fine grade sandpaper in order to reduce the resistance between the pair of electrodes. While the resistance was not measured directly (an extremely low voltage ohmmeter was not available and the phosphene effect of passing a current through the electrodes might have produced undue anxiety in the naive *S*s) the resistance was considered sufficiently low when a calibration eye movement of 7° produced a pen deflection of at least 5 mm. on the eye movement channel with the gain setting used.

Instructions.—The instructions to *S* were to fixate a centrally placed dot in the fixation field upon a "ready" signal, to refrain from moving his eyes until the stimulus appeared, to move his eyes to whatever numeral or numerals he pleased when the stimulus appeared.

The *S* was told to write on an answer sheet after each trial the numerals he had seen in the positions in which they had appeared.

Subjects.—Fourteen naive college students served as *S*s.

RESULTS

The data support the hypothesis that a congruence in direction exists between the initial postexposure eye movement and the more accurate field on trials in which there was both an eye movement and an accuracy difference between the fields. There were more congruent than noncongruent trials ($t = 5.05$, $P < .01$), and this relationship held for 13 of the

14 *S*s (a binomial test giving $P < .01$).

This congruence also appears in each viewing condition individually. Table 1 presents the frequency of occurrence of trials with each combination of eye movement direction and accuracy difference for each viewing condition.

Disregarding perceptual accuracy, in both monocular sets of trials there are twice as many initial eye movements in the direction of the viewing eye than in the opposite direction; i.e., with right eye alone there are more initial eye movements to the right than to the left ($t = 4.12$; $P < .01$) and with left eye alone there are more initial eye movements to the left than to the right ($t = 5.36$; $P < .01$).

Scorable initial postexposure eye movements (in a latency range of from 150 msec. to 1000 msec. after stimulus onset) did not occur on about 30% of the trials. This can, in part, be attributed to the electro-oculograph method used. It was found to be subject to interference from eye blinks and occasional galvanic skin response potentials when such occurred in the latency range chosen.

Perceptual accuracy (defined as correct numeral reported in its correct location) was moderately high, with 52.9% of all reported numerals correct. However, *S*s recorded only 35.8% of all reportable numerals.

DISCUSSION

The results of the present experiment reveal a relation between relative accuracy of perception of numerals on either side of the fixation point and the direction of the initial postexposure eye movement. The conditions of the experiment effectively precluded any visual feedback from eye movement which could lead to differential accuracy on the two sides of the field.

The significance of the present finding lies in the clear indication that other processes than visual feedback are significantly related to accuracy of report of visual material. However, the design of the study does not permit differentiation between the effects of tendencies to eye movements and the effects of possible kinesthetic feedback from the initial eye movement, a task for later empirical determination. Further, one explanation of the findings of a visual nature might lie in the possibility that the clearer perception of numerals on one side of the field, for whatever reason, might lead to an actual eye movement in that direction of relatively short latency. Congruence occured on 85% of the trials in which the latency of initial postexposure eye movements was from 150 to 175 msec. Whether such short latencies are consistent with an alternative explanation in terms of an underlying visual discrimination process is a question for more discriminating empirical determination.

The clear demonstration of an association between accuracy of report and eye movement direction urgently raises the need for clarification of the detailed ways in which tendencies to movement or other relevant variables may affect perception and calls for tests of alternative hypotheses. For instance, one problem is whether the motor tendency directly affects properties of the visual reaction itself or is more intimately related to the process of accurate report.

A secondary positive finding which also invites specific explanation is that initial postexposure eye movements tend to be made to the left when viewing with the left eye and to the right when viewing with the right eye. At a physiological level, the movement of the viewing eye is related to action of the external rectus muscle, and movements are made toward stimuli exciting nasal retina; while at a psychological level a possible relevant fact might be that *S* adjusted the eye aperture occluder and was "aware" that he viewed with a given eye alone, eye movements being in the direction away from the occluded aperture. Again an experimental determination between alternative explanations might give insight into the details of the processes involved.

Regardless of ultimate explanation, the findings of the present experiment indicate the necessity of taking into account direction of initial postexposure eye movements as related to accuracy differences between the fields in tachistoscopic studies, and gives evidence that the presentation of stimuli at exposure times short enough to rule out fixation eye movements does not eliminate the possible effects of eye movements, or tendencies to eye movements, upon the subsequent report.

SUMMARY

The direction of initial postexposure eye movements was studied in a tachistoscopic situation in which a row of numerals appeared across the visual field and no eye movement occurred until the cessation of stimulation. A congruence was found between the direction of the initial eye movement and the side of the visual field more accurately perceived. This finding supports the hypothesis that differential tendencies to eye movement are associated with differential accuracy. A secondary finding was that, in monocular viewing, there were more eye movements to the side of the viewing eye.

REFERENCES

CROVITZ, H. F. Patterns of relative localization of an odd element within a visua- grouping as a function of laterality charl acteristics and tendencies to eye movement- Unpublished doctoral dissertation, Duke. University, 1960.

FORD, A., & LEONARD, J. L. Techniques for recording surface biolectric direct currents. *USN Electron. Lab. Rep.*, 1958, No. 839.

GAFFRON, M. Right and left in pictures. *Art Quart.*, 1950, **13**, 312–331.

HERON, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70**, 38–48.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 27, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 499–504

THE EFFECT OF NONSENSE-SYLLABLE COMPOUND STIMULI ON LATENCY IN A VERBAL PAIRED-ASSOCIATE TASK[1]

BARBARA S. MUSGRAVE[2]

University of Massachusetts

The present experiment was concerned with effects of convergent, associated-nonassociated, and divergent stimulus compounds on latency of responses in a verbal paired-associate (PA) task. The paradigm for convergent compounds is to train *S*s to respond with the same response to each of the stimulus elements presented separately, and then to present the elements in combination. For divergent compounds, the procedure is the traditional conflict paradigm in which *S*s are trained to respond with different responses to each of the stimulus elements, which are then presented together. The paradigm for associated-nonassociated compounds is to train *S*s to respond to a stimulus element and then to present that element accompanied by a new, previously unpresented, element.

According to a fuller review presented elsewhere (Musgrave, 1960), findings with stimulus compounds in relatively simple learning tasks suggest that convergent compounds might facilitate performance, such facilitation being explained in terms of summation of response strengths. On the other hand, findings of inhibitory effects of response conflict suggest that divergent compounds might hinder performance. Evidence concerning the effects of including a novel element in the stimulus are somewhat ambiguous but may be interpreted that performance with associated-nonassociated compounds would be intermediate between that for convergent and that for divergent compounds.

Only Shepard and Fogelsonger (1913), using a PA technique, have investigated the effects on latency and recall of combining into verbal stimulus compounds the single stimuli of experimentally established verbal associations. Unfortunately, none of their experiments was free of one or more methodological shortcomings. The present study, therefore, was designed to refine and extend information on effects of convergent, associated-nonassociated, and divergent compounds in a verbal PA task.

METHOD

Design.—In addition to the experimental conditions using the three different types of compounds, three control conditions were used: Continued Singles, in which a single syllable of the training phase was the stimulus of the test phase; New Singles and New Doubles in which the stimuli of the test phase had not appeared during the training phase. The Continued Singles condition permitted comparisons between compounds and elements composing the compounds. The New Singles and New Doubles were controls for effects of such factors as familiarization with the mode of presentation of the stimuli and experience with the PA technique.

---

[1] This paper is based upon a dissertation submitted to the Department of Psychology of the University of Massachusetts in partial fulfillment of the requirements of the PhD degree, and was supported by a research fellowship, MF-6860-C, from the National Institute of Mental Health, United States Public Health Service. The writer is greatly indebted to A. E. Goss under whose direction this investigation was conducted. Thanks are due also to J. L. Myers for help with the statistical design.

[2] Now at Smith College.

Because written or spoken words are typically sequential, order of the stimulus elements in the sequence may be an important variable, with the element nearest the response the more effective (Howes & Osgood, 1954; Musgrave, 1958). Therefore, order was introduced as a second variable and the nonsense-syllable elements of each of the kinds of compound appeared in both orders. Order of the syllables, when each is the stimulus for the same response, as in the Convergent condition, involves nothing more than counterbalancing for occurrence in the first or second position of the particular syllables used. In both associated-nonassociated and divergent compounds, however, training differences are involved. Consequently, order is not orthogonal to type of compound. For convenience, the two orders used were labeled I and II. Order I refers to compounds in which the syllable which appears first, reading from left to right, is the stimulus whose response was to be considered correct on the test trials. Order II refers to compounds in which this critical stimulus occurs second.

Subjects.—Seventy-two undergraduate women were tested individually, being assigned as they appeared to two groups of 8 *S*s each to serve in the two counterbalanced Convergent conditions, to two groups of 4 *S*s each to serve in the two counterbalanced New Doubles conditions, and to six groups of 8 *S*s each to serve in the remaining six conditions. Forty-eight additional *S*s were tested but discarded because of failure to meet one or more of three criteria of performance on the single-syllable stimuli of the training lists. These criteria are described in the section on procedure.

Apparatus.—The paired associates were presented by means of a modified Hull memory drum. With the opening of a left-hand shutter, the stimulus member of each pair was presented alone for 2 sec., then in addition to the stimulus, with the opening of a right-hand shutter, the response member was presented for 2 sec. The interpair interval was 4 sec. Latency of responses was obtained by means of a Gerbrand's electronic voice key and a Hunter Klockounter. The sound of the first letter of the spelled response, acting through the voice key, stopped the Klockounter.

Paired associates.—Forty nonsense syllables, selected from Glaze's (1928) list of syllables with high association values (100% and 93%) and of low similarity to each other, were used. In order that test materials for all conditions would contain the same syllables, differing training lists were used in different conditions.

The training materials included three kinds of items: (*a*) the basic list of single-syllable stimulus and response pairs, (*b*) pairs of associated-nonassociated stimuli and single-syllable responses, and (*c*) either single-syllable or associated-nonassociated stimuli alone, without their appropriate responses. The basic list consisted of four pairs made from four single-syllable stimuli and two single-syllable responses. In the associated-nonassociated pairs, the four stimuli in the basic list were each accompanied by a second syllable which had not previously appeared and did not appear again. These associated-nonassociated pairs were used to prevent formation of a set-for-single-stimuli by providing experience with compound stimuli (Andreas, 1958). Also, they served to accustom *S* to respond to stimulus compounds and to respond with only one or the other of the two responses of the learning task. The third type of item—the stimulus presented alone—was used to accustom *S* to a response period of 4 sec. rather than 2 sec. This lengthened response period seemed desirable to use on the first test trial following training because preliminary studies had shown that 2 sec. was insufficient time for anticipatory responses under Divergent, New Doubles, and New Singles conditions.

The training materials consisted of 116 items, divided into four lists of 24 items each, and a fifth list which began with 20 training items and ended with the items for the four test trials. Before each *S* arrived all five lists for the appropriate condition were arrayed on the drum. After the first list of 24 was presented, the drum was moved along its axle and rotated to the starting position for presentation of the next list of 24 and so on.

The items for the four test trials consisted of a single stimulus-response pair presented four successive times. The stimuli for the New Singles and Continued Singles were single syllables. For the remaining six conditions, the stimuli were two-syllable compounds. On Trial 1, the stimulus was followed by a blank. On the remaining three trials, the stimulus was followed by responses designated correct by *E*.

Procedure.—The order in which the conditions were presented was used to assign *S*s to groups. The order was random with the restriction that every condition should be presented once before any condition was presented twice and so on. Within this restriction, the 54 different counterbalanced combinations of training and test materials (72 combinations in all) were assigned to a schedule sheet with the aid of a table of random numbers.

Each *S* was seated in front of the apparatus, asked to hold the microphone close to her mouth and to speak the word "Hello" when each of the shutters opened to reveal the stimulus and then the response. This procedure was explained to *S* as necessary for adjusting the apparatus for her individual voice, and permitted coaching her in speaking loudly and clearly and in responding twice on every trial, once before and once after the opening of the second, or response, shutter. The *E* then sat out of sight behind the apparatus and read the instructions. In addition to general information concerning the PA task, instructions included the specific information that there would be only two response syllables. What these syllables would be was told to *S* six times. The *S* then practiced with the training materials. The final list, as mentioned above, ended with four successive presentations of the particular compound stimulus or, in the case of the New Singles and Continued Singles conditions, the single-syllable stimulus which constituted the test stimulus. No new instructions or other comments preceded the presentation of the test trials. On the first occurrence of the test stimulus, no response member was presented. Thus, on this trial, as the test stimulus was presented alone for 4 sec., it was possible to measure latencies up to 4 sec. rather than up to 2 sec. On Trials 2, 3, and 4 of the test stimulus, the correct response member was exposed for 2 sec. Thus, Test Trials 1 and 2 were, in effect, recall trials, while Trials 3 and 4, following the initial presentation of the correct response on Trial 2, were learning trials.

Following completion of each *S*'s participation, her performance was evaluated on the basis of three criteria: (*a*) correct responses on the last two presentations of each of the critical training syllables, i.e., the syllables which in the Convergent and Divergent conditions were paired together as compound stimuli in the test trials, or the corresponding stimuli in the training lists of the other conditions; (*b*) response latencies within .300 sec. of each other on the last occurrence of the two critical syllables; (*c*) an equal number of correct responses, plus or minus one response, for the 11 times each syllable appeared in the last three lists. The first of these criteria was included to insure that *S*s had actually learned the correct responses to the critical syllables before entering the test phase.

The second and third criteria were of particular importance in the Divergent (conflict) condition, as they were designed to insure that responses to the two syllables were at equal or near equal strength, thus guarding against the possibility that choice of responses reflected nothing more than differential strengths. To equate selection of *S*s through conditions, these two criteria were applied not only to the Divergent but to all other conditions as well. In connection with the third criterion, it was assumed from preliminary studies that the numbers of correct responses would be fairly stable across conditions. However, it was decided that any numbers of correct responses which deviated from the numbers of correct responses for the majority of *S*s, even if equal between the two critical syllables, would be sufficient cause for replacing an *S*. If the performance of any *S* did not meet these criteria, her protocol was discarded and the next *S* to appear was assigned to the same condition.

RESULTS

Training Trials

The single syllables which were to appear in the test compounds appeared in the last three training trials 11 times. The *S*s of all groups responded correctly to these syllables either 10 or 11 times.

Latencies on the last training trial, the criterion trial, for each of the two single syllables which were to be combined in the test compounds were averaged. Means and *SD*s of these averaged latencies for all eight conditions are presented in Table 1. The latencies of both criterion and test trials exhibited marked heterogeneity of variance. Accordingly, for the analysis of variance, these data were transformed into their reciprocals (speed of response evocation). The analysis of variance of the criterion-trial data indicated that there were no significant differences among conditions at the end of the training trials ($F < 1$; $df = 7/64$).

TABLE 1

MEANS AND *SD*s OF LATENCIES ON THE CRITERION TRAINING TRIAL AND FOUR TEST TRIALS FOR ALL CONDITIONS

Condition	Training		Test							
	Criterion Trial		Trial 1		Trial 2		Trial 3		Trial 4	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Continued Singles (Cont. S.)	0.958	.090	0.902	.123	0.960	.212	0.904	.114	0.917	.145
Convergent (Conv.)	0.906	.100	1.113	.205	1.064	.212	1.000	.179	1.001	.263
Asso.-Nonasso. I (A-NA I)	0.911	.071	1.196	.237	1.099	.249	1.028	.173	0.980	.167
Asso.-Nonasso. II (A-NA II)	0.961	.179	1.233	.279	1.088	.167	1.006	.164	1.065	.219
Divergent I (Div. I)	0.981	.131	1.236	.164	1.635	.700	1.267	.363	1.121	.198
Divergent II (Div. II)	0.941	.188	1.240	.501	1.162	.255	1.451	.625	1.129	.278
New Singles (New S.)	0.928	.084	1.969	.852	1.620	.779	1.134	.400	0.990	.155
New Doubles (New D.)	1.006	.095	2.680	1.099	1.812	.938	1.162	.819	1.082	.739

Test Trials

Latencies on all four test trials.—Means and *SD*s of latencies for all eight conditions for each of the four test trials are presented in Table 1. The *F*s for Conditions, Trials, and Trials × Conditions were significant at less than .001 (Table 2). Because of the significant Trials × Conditions interaction, separate analyses of variance were carried out for each trial. Differences among Conditions were significant for Trials 1 ($P < .001$), 2 ($P < .001$), and 3 ($P < .01$) but were not significant for Trial 4.

The results of a Duncan multiple range test of the significances of differences between all pairs of means of speeds (reciprocals of latencies) on each of the four trials are summarized in Table 2.

An analysis of variance for each condition separately assessed the effects of order and of trials both for the Divergent and Associated-Nonassociated conditions in which order of syllables was based on training differences and for the Convergent and New Doubles conditions in which order reflected only a counterbalancing of particular syllables. In no case was order a significant variable ($F < 1$ in all cases). However, across trials, significant decreases in latencies occurred under Associated-Nonasso-

TABLE 2

SUMMARY OF DUNCAN RANGE TESTS

Cond.	Cont. S.	Conv.	Div. II	A-NA I	A-NA II	Div. I
Trial 1						
Conv.	.05	—	—	—	—	—
Div. II	—	—	—	—	—	—
A-NA I	.05	—	—	—	—	—
A-NA II	.05	—	—	—	—	—
Div. I	.05	—	—	—	—	—
New S.	.01	.01	.01	.05	.05	.05
New D.	.01	.01	.01	.01	.01	.01
Trial 2						
New S.	.01	.01		.05	.05	
Div. I	.01	.01		.05	.05	
New D.	.01	.01		.01	.05	
Trial 3						
New D.	.05	—		—	—	
Div. I	.01	.05		—	—	
Div. II	.01	.01		.05	.05	

Note.—Conditions in the margins both from top to bottom and left to right of each table are listed in the order of magnitude of means for each trial, except for a reverse positioning of A-NA II and A-NA I in the table for Trial 3. A-NA II and A-NA are not reversed on Trial 2 due to a shift effected in the direction of the means by the transformation from latencies to speeds.

ciated, New Doubles, and New Singles conditions.

Number correct on recall trials.—For the Convergent, Associated-Nonassociated, and Continued Singles conditions, which had been trained with only one response syllable designated as correct, all *S*s responded correctly on Test Trials 1 and 2. For the Divergent, New Doubles, and New Singles conditions *either* of the two trained response syllables was scored as correct on the recall trials. On Trial 2 1 *S* in Divergent I responded after the opening of the second shutter had revealed the correct response and, hence, her response was marked incorrect. In the New Doubles condition, 1 *S* failed to respond on Trial 1, and a second *S* made no response on either recall trial. One *S* in the New Singles condition made no response on Trial 1 and waited to read the response syllable on Trial 2; these responses were scored incorrect.

Number correct on learning trials.—After exposure of the stimulus on Trial 2, the correct response syllable had been exposed. Thereafter, on Trials 3 and 4, only this syllable was scored as correct. Only *S*s in the Divergent condition failed to respond correctly. Three *S*s responded with the incorrect syllable on Trial 3, 1 responded with the incorrect syllable on both Trials 3 and 4, and another failed to respond on Trial 3 before the second shutter opened.

DISCUSSION

The expected rank order of increasing latencies for the experimental conditions of Convergent, Associated-Nonassociated, and Divergent was confirmed. However, the Convergent condition did not facilitate performance relative to the Continued Singles. That facilitation was possible is indicated by the fact that latencies for the Continued Singles condition were shorter on all four test trials than those for any condition on the last training trial. In contrast, latencies for the Convergent condition were longer than those for any condition on the last training trial, and were longer than those for the Continued Singles condition throughout the four test trials. Either the convergent associations did not summate, or the effects of summation were outweighed by factors impairing performance. For example, it may have taken longer to look at and perhaps to read two-syllable stimuli. Or the poorer performance may have been due to dissimilarity between presentation of one nonsense syllable alone and presentation of that syllable simultaneously with another nonsense syllable. The *S*s had had experience with each of the test syllables presented with another syllable but their training had not included presentation of the test syllable with another syllable to which a response had been associated experimentally.

The Associated-Nonassociated conditions resulted in consistently poorer performance than the Convergent condition, but the differences were slighter than expected. Possibly presentation of Associated-Nonassociated compounds during training had accustomed *S*s to quick selection of the associated syllable and immediate disregard of the nonassociated syllable. In the introduction of a novel stimulus element, there are two sources of novelty. The introduction may be novel, the element may be novel, or both. With the procedure followed here, the introduction of a novel element was no longer completely novel, nor was doubleness as such. However, the nonassociated element of the first test stimulus was novel in that it had not been presented previously in the experimental situation. Presumably, then, the novelty of this element was chiefly responsible for the differences between the Associated-Nonassociated conditions and the Continued Singles and Convergent conditions.

Latencies on Trial 1 did not fall as low for the New Singles condition as for the New Doubles, and the slope of the improvement curve was steeper for the New Singles. These results indicate the desirability of enlarging available information concerning the difficulty of multi-element stimuli, possible interaction of number of stimulus elements with length of list, and so on.

Results for the Divergent conditions differed sharply and in several ways from those for the other five conditions. In general, what is thought of as typical conflict behavior was exhibited. In addition to long latencies, responses of *S*s in the Divergent condition showed vacillation, providing instances not only of response blends but also of response alternation from trial to trial even after the correct response had been presented.

SUMMARY

The present study was concerned with the effects on response latency of convergent, divergent, and associated-nonassociated nonsense-syllable compounds. Order of the syllables in the divergent and associated-nonassociated compounds was a second, nonorthogonal variable. Continued Singles, New Singles, and New Doubles conditions served as controls.

On the first of four test trials, all experimental compound conditions were slower than Continued Singles and faster than the New Singles and the New Doubles conditions. Although differences among the experimental compound conditions were slight, the rank order of increasing latencies of the Convergent, Associated-Nonassociated, and Divergent conditions was that which had been predicted.

On the remaining three test trials, Continued Singles did not improve, suggesting that asymptotic latencies had been reached. In contrast, latencies for the Convergent condition and for the two Associated-Nonassociated conditions decreased moderately, while those for the New Singles and New Doubles conditions decreased markedly. Only the two Divergent conditions did not exhibit either asymptotic latencies or reasonably steady improvement.

REFERENCES

ANDREAS, B. G. Motor conflict as a function of motivation and amount of training. *J. exp. Psychol.*, 1958, **55**, 173–178.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

HOWES, D., & OSGOOD, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 1954, **67**, 241–258.

MUSGRAVE, B. S. Context effects on word associations using one-word, two-word, and three-word stimuli. Paper read at Eastern Psychological Association, Philadelphia, April 1958.

MUSGRAVE, B. S. Response speed, recall, and learning with nonsense-syllable compound stimuli. Unpublished doctoral dissertation, University of Massachusetts, 1960.

SHEPARD, J. F., & FOGELSONGER, H. M. Studies in association and inhibition. *Psychol. Rev.*, 1913, **20**, 290–311.

(Received March 27, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 505–509

COMBINATION OF DRIVE AND DRUG EFFECTS

JOSEPH MENDELSON AND DALBIR BINDRA

McGill University

Does the effect of a drug on a trained response depend upon the current drive level of *S*? In other words, is the performance of a given response more vulnerable to the effects of a drug when the response is tested at a high drive level than when it is tested at a low drive level? An answer to this simple empirical question is likely to contribute to an understanding of the way in which drive (*D*) combines with other variables in determining response output, and might clarify the results of the studies (see Hulicka, 1960, for review) of the interaction of *D* with habit strength (*H*) and reward (*K*).

In the present investigation, the combinations of each of several levels of drive (thirst) with two drugs were studied. Variations in thirst drive were obtained by "prefeeding" *S*s different amounts of water before testing. The drugs employed were Chlorpromazine, a mild depressant ("tranquilizer"), and methylphenidate, an amphetamine-like stimulant. Though, like other drugs described as stimulants, methylphenidate, in medium doses of from 2 to 10 mg/kg, increases general activity (Bindra & Baran, 1959), we have repeatedly observed that it decreases the rate of a water-, food- or saccharin-rewarded lever pressing response; in this respect methylphenidate resembles amphetamine, which is also known to decrease such responding under certain conditions (e.g., Hearst, 1961; Miller, 1956; Owen, 1960). By using doses of Chlorpromazine and methylphenidate that would have roughly equal decremental effects, we hoped to study the interactions of drive with these drugs at comparable levels of response output.

METHOD

Subjects.—Twenty-eight naive male hooded rats, about 90 days old (weighing between 170 and 205 gm.) at the beginning of the experiment, were used as *S*s. They were housed in small metal cages, 2 to a cage. Food was available to them at all times except during the experimental sessions. The *S*s were placed on a 23-hr. schedule of water deprivation the day after their arrival at the laboratory, and they remained on this schedule for the duration of the experiment. They were allowed access to water in their home cages for 40 min. daily, at the conclusion of the experimental session.

Apparatus.—The apparatus consisted of four identical lever boxes, made of plywood and measuring 7 in. wide, 8½ in. deep, and 11½ in. high. All the boxes were painted gray on the inside and were covered by a wire mesh roof. Each was equipped with a water-delivery mechanism and an automatic counter. A lever, measuring 4 × ¾ in., was mounted 3 in. above the floor on one of the smaller walls of each box. A tray for receiving water was located on the same wall, 2½ in. from the center of the lever and 2 in. above the floor. Each lever press delivered .02–.04 cc of water into the tray. The minimum weight required to depress the lever was 44 gm.

Procedure.—After 1 day of water deprivation, each *S* was placed in a lever box for 10 min. The trays were filled with water; by the end of the period most *S*s had begun to drink the water. A 13-day training period was initiated on Day 2. On Days 2–9 each *S* was placed in a lever box for 15 min.; on Days 10–14, for 10 min. A given *S* was always placed in the same one of the four lever boxes. At the end of the training period the 4 *S*s with the lowest lever pressing rates were discarded, leaving 24 for the tests.

Variations in thirst drive were produced by allowing each *S* to drink a predetermined amount of water during a 10-min. period beginning about 40 min. before the daily test

TABLE 1

SUMMARY OF EXPERIMENTAL PROCEDURE

Stage	"Predrinking" in cc			Drug Condition	
	High Drive	Medium Drive	Low Drive	Drug Injection	Water Injection
1	0.0	2.5	5.0	1.5 mg/kg Chlorpromazine	0.3cc
2	2.5	7.5	10.0	1.5 mg/kg Chlorpromazine	0.3cc
3	0.5	4.0	8.0	4.0 mg/kg methylphenidate	0.75cc

session. Each *S* was placed with its allotted quantity of water in a box measuring 7 × 8 in. Except when the highest amount (10 cc) of water was given, all *S*s drank all their allotted water in the given 10-min. period. When *S* did not finish the allotted water, it was placed in its home cage and given an additional 10-min. period to finish the water; at the conclusion of this period there was never more than 1 cc of water left over.

Immediately after water consumption, that is, about 30 min. before testing, each *S* was weighed and injected intraperitoneally with an appropriate amount of distilled water (control), Chlorpromazine, or methylphenidate. The test sessions were conducted in three stages, each consisting of six sessions, one per day, with one practice session in between the stages. During each stage, each *S* was tested twice under each of the three drive levels. The two test sessions at each drive level were conducted on consecutive days; one of the tests followed a drug injection, while the other followed a control injection. The order of the three drive levels and the two types of injection (drug vs. control) was varied so that 2 *S*s were tested in each of the 12 possible permutations of drive level and type of injection. Thus, each of the 24 *S*s was tested under the six conditions (3 drive levels × 2 drug conditions) in each of the three stages of the experiment. The exact conditions of drive and drug employed in each stage are summarized in Table 1. It may be seen that, in Stages 1 and 2, *S*s were tested under the influence of 1.5 mg/kg Chlorpromazine at five different drive levels; in Stage 3, effects of methylphenidate at three drive levels were determined.

RESULTS

The mean numbers of responses made in the 10-min. test sessions under the different conditions of drive and drug are shown in Fig. 1 and 2. The higher scores obtained in Stages 2 and 3 can be accounted for by the fact that *S*s had not been trained to an asymptote before the initiation of the test sessions, so that their perform-

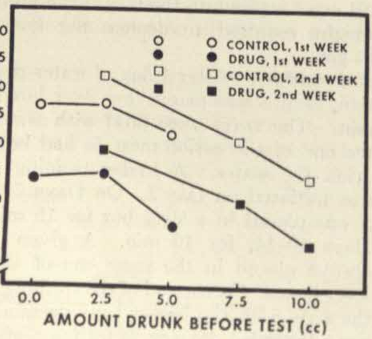


FIG. 1. Number of lever presses as a function of five drive levels (high to low) following control (water) and drug (Chlorpromazine, 1.5 mg/kg) injections.

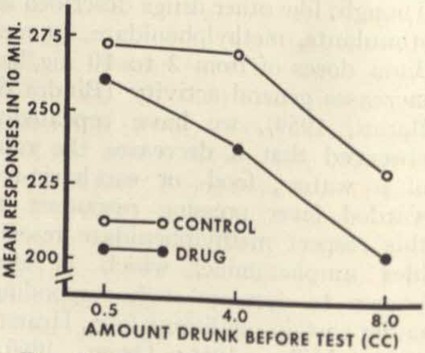


FIG. 2. Number of lever presses as a function of three drive levels (high to low) following control (water) and drug (methylphenidate, 4 mg/kg) injections.

TABLE 2

VALUES OF t FOR DIFFERENCES BETWEEN MEANS AT DIFFERENT DRIVE LEVELS

Stage	Control Scores			Drug Scores		
	D_L-D_M	D_M-D_H	D_L-D_H	D_L-D_M	D_M-D_H	D_L-D_H
1	2.72*	0.04	2.39*	3.35**	0.46	2.34*
2	3.93**	4.03**	9.29**	2.61*	2.50*	5.20**
3	3.54**	0.31	3.85**	2.71*	1.48	4.30**

* $P < .05$.
** $P < .01$.

ance was still improving during the test sessions. Three features of the curves in these figures should be noted. First, in the control, water-injection, conditions, the higher the drive level (i.e., the less the amount of "predrinking") the greater were *S*s' response rates. Two-tailed t tests for differences between correlated means indicated that most of the differences between the means at different drive levels were significant (see Table 2); the only nonsignificant differences were those between medium and high drive levels in Stages 1 and 3. Second, the level of performance in the drug conditions was also different at the different drive levels; the relation between drive level and response rate was the same as that exhibited under control conditions—the higher the drive level, the greater the rate of response. Again most of the differences were significant (see Table 2); as in the case of control scores, the differences between the scores at the medium and high drive levels in Stages 1 and 3 were not significant. Third, both drugs led to a decrease in the number of lever presses; the differences between the mean drug and mean control performances were significant at each drive level ($P < .05$), except under the high drive methylphenidate condition.

To determine whether the effects of the drugs on response rate varied with drive level, two measures of response decrement were calculated for each *S*: (*a*) the absolute difference between drug and control scores, and (*b*) the ratio of drug score to control score (D/C). The latter measure was used to minimize the effects of the absolute level of performance on response decrement. Using the absolute measure, the t test for correlated means indicated no significant differences between the means of absolute decrement in response scores at the different drive levels; this was true for both drugs. Using the ratio measure (Table 3) it was also found that the mean ratio of drug to control performance was independent of the drive level at the time of test. In the case of Chlorpromazine, the ratios at the lowest drive level were somewhat smaller than those at the two higher drive levels, but this apparent trend is not reliable.

DISCUSSION

Under the conditions of the present experiment, mild doses of Chlorpromazine and methylphenidate decreased the rate of lever pressing for water. The decremental effect of methylphenidate requires a word of comment. Even a cursory glance at the results of studies of the effects of stimulants on a lever

TABLE 3

MEANS AND *SD*s OF RATIOS OF DRUG TO CONTROL PERFORMANCE AT DIFFERENT DRIVE LEVELS

Stage	Drive Level					
	Low		Medium		High	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	.58	.31	.73	.28	.75	.22
2	.60	.36	.79	.30	.76	.28
3	.94	.28	.89	.29	.97	.19

pressing response (e.g., Hearst, 1961; Miller, 1956; Owen, 1960; Skinner & Heron, 1937; Teitelbaum & Derks, 1958) shows these effects to be quite variable; incremental effects are found together with decremental effects, sometimes in the same experiment. We believe that these discrepancies have arisen mainly from three types of differences in procedure: (*a*) Dose level: in general, the higher the dose the more likely it is that the stimulant will produce a decremental effect. (*b*) Schedule of reinforcement: continuous reinforcement schedules are more likely to yield decremental effects; schedules, such as partial and extinction, under which the animal makes many multiple responses (lever presses without entering the goal area) are likely to yield incremental effects. (*c*) Distance between the lever and the reward area: in general, the greater this distance, the more likely it is that the stimulant will decrease response rate. (It may be that the crucial factor here is the size of the lever box rather than the distance between the lever and the reward area). Using the specifications of the procedure employed in the present experiment, we have consistently obtained decremental effects with from 2 to 10 mg/kg of methylphenidate. As methylphenidate, like amphetamine, has an anorexic effect (Karczmar & Howard, 1959), it cannot be said whether the obtained decrease in lever pressing rate arises from anorexia or from interference caused by increased general activity; however, we are inclined to attribute greater importance to the latter factor (Bindra, 1961).

The decremental effects of both Chlorpromazine and methylphenidate, as measured by the absolute and proportionate decrement in response scores in the present study, were fairly constant over a wide range of drive levels. For Chlorpromazine, the effects were constant over the entire range of drive levels used (see Fig. 1 and Table 3). For methylphenidate, although the effects at different drive levels were not significantly different from each other, the absolute decrement in response rate was significantly different from zero at the two lowest drive levels, but not at the highest drive level (see Fig. 2 and Table 3). Our casual observations suggested one possible basis for the difference between Chlorpromazine-drive and methylphenidate-drive interactions. It seemed to us that while the decrement in the response rate under the influence of Chlorpromazine resulted from over-all slowing of *S*'s movements, the methylphenidate-induced decrement was associated with an increase in general activity, mainly repeated sniffing of different parts of the apparatus. Whether this difference is in any way related to the observed differences in interaction effects is a question that we are now investigating.

Payne (1958), in extending Hullian theory to account for response decrements resulting from drugs, concluded that "manipulations resulting in higher reaction potential will, under constant dosage, attenuate the degrading effect of a drug" (p. 342). According to Payne's analysis, one would expect the depressing effects of drugs to be less when the response is tested at high drive than when it is tested at a low drive level. The present results show this to be true only in the case of methylphenidate. Apparently, the simple relation suggested by Payne has limited applicability. The factors that determine the exact manner in which various drive and drug effects combine remain to be determined.

SUMMARY

Is the effect of the injection of a given drug on a response dependent upon the drive level at which the response is tested? The effects of injections of Chlorpromazine (1.5 mg/kg) and methylphenidate (4 mg/kg) on a water rewarded lever pressing response were studied at several levels of thirst. Within three phases of the experiment all *S*s were tested, in a balanced order, under each of the six conditions (3 drive levels × 2 drug conditions).

The following results were found: (*a*) The higher the drive level, the greater the response rate in both the control and the drug conditions. (*b*) Both Chlorpromazine and methylphenidate decreased response rate, though the former drug is a mild depressant

("tranquilizer") and the latter a stimulant. (*c*) There were no significant differences between the mean absolute, or relative, decrements in response score produced by either of the drugs at the different drive levels. (*d*) Methylphenidate failed to produce any marked decrement at the highest drive level; thus the response was less vulnerable to the effects of this drug at a high drive level than at low drive levels.

The decremental effects on the response of stimulants, such as methylphenidate and amphetamine are discussed. Some suggestions are presented regarding the basis of the difference in methylphenidate-drive and Chlorpromazine-drive combinations.

REFERENCES

BINDRA, D. Components of general activity and the analysis of behavior. *Psychol. Rev.*, 1961, **68**, 205–215.

BINDRA, D., & BARAN, D. Effects of methylphenidylacetate and Chlorpromazine on certain components of general activity. *J. exp. Anal. Behav.*, 1959, **2**, 343–350.

HEARST, E. Effects of d-amphetamine on behavior reinforced by food and water. *Psychol. Rep.*, 1961, **8**, 301–309.

HULICKA, I. M. Combination of drive and incentive. *Quart. J. exp. Psychol.*, 1960, **12**, 185–189.

KARCZMAR, A. G., & HOWARD, J. H. Anorexigenic action of methylphenidate (Ritalin) and Pipradrol (Meratran). *Proc. Soc. Exp. Biol. Med.*, 1959, **102**, 163–167.

MILLER, N. E. Effects of drugs on motivation: The value of using a variety of measures. *Ann. N. Y. Acad. Sci.*, 1956, **65**, 318–333.

OWEN, J. E., JR. The influence of dl-, d-, and l-amphetamine and d-methamphetamine on a fixed-ratio schedule. *J. exp. Anal. Behav.*, 1960, **3**, 293–310.

PAYNE, R. B. An extension of Hullian theory to response decrements resulting from drugs. *J. exp. Psychol.*, 1958, **55**, 342–346.

SKINNER, B. F., & HERON, W. T. Effects of caffeine and benzedrine upon conditioning and extinction. *Psychol. Rec.*, 1937, **1**, 340–346.

TEITELBAUM, P., & DERKS, P. The effect of amphetamine on forced drinking in the rat. *J. comp. physiol. Psychol.*, 1958, **51**, 801–810.

(Received April 3, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 510–513

ANAGRAM SOLUTION TIMES: A FUNCTION OF WORD TRANSITION PROBABILITIES

M. S. MAYZNER AND M. E. TRESSELT
University of Southern California *New York University*

Mayzner and Tresselt (1958, 1959) demonstrated anagram solution times to be a function of three variables: (*a*) anagram letter order, (*b*) word frequency, and (*c*) transition probability. The first variable, anagram letter order, refers to the rearrangement of the letters from their correct position in the word, while the second variable, word frequency, refers to the frequency of occurrence in English of the anagram solution as estimated from Thorndike-Lorge (1944) counts. The third variable, transition probability, refers to the probability or frequency with which given letters follow or precede other letters and digram frequency tables are available (Pratt, 1942; Underwood & Schulz, 1960) which permit the relative assignment of a numerical value to this frequency. For this variable it was found that anagrams whose summed transition probability or digram frequency totals are low yield significantly faster solution times than anagrams whose summed transition probability totals are high.

The present study examines the effect of word transition probability on anagram solution times, by selecting anagrams whose solutions are words yielding either relatively low summed digram frequency totals or relatively high summed digram frequency totals. It is predicted that anagrams whose word solutions have high transitions or digram totals will be solved significantly faster than anagrams whose word solutions have low transition totals, since it is assumed that *S*'s successive rearrangements of the letters of the anagram are more likely initially, at least, to correspond to sequential letter pairs with high transition probability or digram frequencies than low (e.g., the letters H and E are more likely to be arranged as HE than EH) and therefore more likely match words with high transition probability totals than words with low transition totals.

METHOD

Subjects.—The *S*s were 25 students from introductory psychology courses at Barnard College, Columbia University.

Stimulus material.—The stimulus material consisted of 18 five-letter anagrams typed in caps on white 3 × 5 in. cards. The 18 words from which the anagrams were constructed were selected from the Thorndike-Lorge (1944) word list on the basis of two criteria: (*a*) all words occur 100 times or over per million; (*b*) 9 words have low summed transition probability or digram frequency totals and 9 words have high summed transition totals. Summed transition totals for each word were obtained by adding the digram frequency values, as given by Underwood and Schulz (1960), for the four sequential letter pairs in each word. For example, with the word JUDGE, the four digram frequency values for the letter pairs JU, UD, DG, and GE are 30, 29, 18, and 201, respectively, yielding a summed digram frequency total for the word of 278, while with the word MONTH, the letter pairs MO, ON, NT, and TH yield values of 204, 786, 511, and 1795, respectively, for a summed total value of 3296.

All words and anagrams employed in the study may be found in Table 2. Digram frequency totals for the words range from 278 to 938 for the low transition probability (TP) words with a mean TP value of 725, and from 1647 to 3296 for the high-TP words with a mean TP value of 2190. Since it has been shown that the summed TP total of anagram affects solution time (Mayzner & Tresset

1959), the anagram TP totals were made essentially equivalent for the low- and high-TP words, with a mean anagram TP value of 159 associated with the low-TP words and a mean anagram TP value of 161 associated with the high-TP words. The "similarity" of anagram to word letter order also was minimized in all 18 anagrams. Thus, anagram letter order, word frequency, and anagram TP totals were made essentially equivalent for all 18 words, allowing only the variable under investigation, namely, word transition probability to vary from low to high values.

Procedure.—Each of the 25 *S*s received the 18 anagrams in a different random order, with the constraint that the first, middle, and last 6 anagrams, each contain 3 anagrams whose solutions were high-TP words and 3 anagrams whose solutions were low-TP words. All *S*s, who were tested individually, were told that a five-letter anagram is simply five scrambled letters which, when properly rearranged will make a single five-letter word. After being shown an example, they were informed that they would have 18 anagrams to solve and if they could not find a solution within 2 min. the correct solution would be given to them before going on to the next anagram. Solutions were given by *S* verbally (i.e., *S* had no paper or pencil) and solution times were recorded with a stop watch.

RESULTS AND DISCUSSION

Available for analysis were 18 time scores for each *S*, with each score having a possible range from 1–120 sec. Since each *S* received nine high-TP words and nine low-TP words, the medians of each of these two sets of nine scores for each *S* were used as the criterion measures for statistical evaluation. Medians were employed rather than means because of skewed distributions and the artificial ceiling imposed by the 120-sec. time limit per anagram. The means, *SD*s, medians, and ranges of these criterion measures are presented in Table 1, and the results indicate that anagrams whose word solutions have high-TP totals are solved significantly faster than anagrams whose word solutions have low-TP totals, with the correlated t yielding a value of 2.70 ($df = 24$; $P < .02$). If the sign test is applied to the 25 pairs of scores, 17 pairs are in the predicted direction, 5 pairs opposed, and there are two ties ($P < .01$).

An analysis also was made for each of the 18 individual words. These

TABLE 1

ANAGRAM SOLUTION TIMES (IN SEC.) FOR HIGH AND LOW TRANSITION PROBABILITY CONDITIONS (BASED ON MEDIANS)

Transition Probability	Mean	*SD*	Median	Range
High	14.40	10.95	12.00	3–47
Low	28.00	30.03	16.00	3–120

TABLE 2

ANAGRAM SOLUTION TIMES (IN SEC.) FOR INDIVIDUAL WORDS WITH HIGH AND LOW TRANSITION PROBABILITY TOTALS

High-TP Words	Anagrams	Median	Range	Low-TP Words	Anagrams	Median	Range
WATER	AEWTR	3	1–56	JUDGE	EGUJD	3	1–45
HOUSE	EUOHS	6	1–120	VOICE	EOCVI	4	1–66
DRINK	NRDKI	7	1–83	GUIDE	IUEGD	7	2–120
PAINT	IPTNA	13	2–120	FRUIT	IUFTR	15	1–120
MONTH	OHTNM	13	3–120	MUSIC	IUMCS	27	2–120
HUMAN	MHNUA	15	2–120	VALUE	AEUVL	46	2–120
POUND	UODNP	17	2–120	GUARD	AUGDR	49	1–120
YOUTH	OYHTU	18	1–120	UNCLE	EUCNL	72	1–120
POWER	OEWRP	22	2–94	SUGAR	SRGUA	83	2–120

results are presented in Table 2, which gives the medians and ranges of solution times based on the performance of the 25 *S*s. While five of the low-TP words have longer median solution times (i.e., 27, 46, 49, 72, and 83 sec.) than any of the high-TP words, whose median solution times range from 3 to 22 sec., four low-TP words fall within this range (i.e., 3, 4, 7, and 15 sec.).

Since a previous study by Jenkins and Mayzner (1961) found a significant reduction in anagram solution times over the course of 10 problems, a similar analysis was performed on the present data. No systematic change in solution times was found over the 18 problems. The discrepancy between the two studies is probably a function of the use of a restricted set of anagram letter orders in the previous study, while in the present study each anagram employed a different letter order.

In previous studies Mayzner and Tresselt (1958, 1959) suggested that an anagram constitutes a problem whose stimulus properties tend to evoke a variety of implicit responses on the part of *S*, and letter order, word frequency, and anagram transition probability were related to the way in which these implicit responses were produced. By conceptualizing anagram problem solving in this way, one is, in fact, employing an S-R mediational model of the type being extensively employed by Kendler and his associates (Kendler & D'Amato, 1955; Kendler & Karasik, 1958; Kendler & Mayzner, 1956) in treating concept formation processes. The present study extends this mediational analysis to include the effect of word transition probability on anagram solution times and the results indicate that the model is successful in making predictions regarding the effect of this variable.

While the four variables so far studied appear to be critical factors governing the type of implicit responses produced, they apparently far from exhaust the variance associated with anagram solution times. This is most evident in the wide range of solution times, from 3 to 83 sec., found with the low-TP words. The words GUIDE and GUARD provide an example of this unaccounted for variance, in that the solution time of GUARD (49 sec.) is seven times that of GUIDE (7 sec.), even though four variables have been held constant and three of the five letters are the same in both words. Such findings suggest that further research is required to uncover the variables involved in the production of implicit responses to anagram problems.

SUMMARY

A study was made of the effect of word transition probability on anagram solution times. Word transition probability was varied by selecting words with either relatively high or low summed digram frequency totals, while holding constant anagram letter order, word frequency, and anagram transition probability variables. The results, predicted on the basis of an S-R mediational model, showed that anagrams whose word solutions have high transition probability totals are solved significantly faster than anagrams whose word solutions have low transition totals.

REFERENCES

JENKINS, H. M., & MAYZNER, M. S. The time-pattern of switching among alternatives in problem solving. *Psychol. Rep.*, 1961, **8**, 419–430.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

KENDLER, H. H., & KARASIK, A. D. Concept formation as a function of competition between response produced cues. *J. exp. Psychol.*, 1958, **55**, 278–283.

KENDLER, H. H., & MAYZNER, M. S. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56**, 376–379.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47**, 117–125.

PRATT, F. *Secret and urgent.* New York: Blue Ribbon, 1942.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Bureau of Publications, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received April 7, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 514–518

TRANSITION PROBABILITY EFFECTS IN ANAGRAM PROBLEM SOLVING[1]

HARRY BEILIN AND RHEBA HORN
Brooklyn College

It has been proposed that if certain gestalt assumptions about perceptual and cognitive organization are true, meaningful letter groups such as words should resist reorganization more than meaningless letter aggregates when these are the materials of anagram problem solving experiments (Hollingworth 1935, 1938). Hollingworth and others (Nissenson & Sargent, 1941; Sargent, 1940), however, have found words to be no different from nonsense letter arrangements in resisting alteration. Up to the time of Johnson's (1955) review only Devnich's (1937) findings suggested that a difference exists. This was sufficient for Johnson to consider the question still open.

In the main, the technique used in these experiments required that *S* discover the word (e.g., KANGAROO) that could be made from either related or unrelated word groups (e.g., AGO, KORAN) in contrast with meaningless letter arrangements (e.g., OAG, KRNOA). More recently, Mayzner and Tresselt (1958, 1959) have tested the aforementioned thesis employing meaningless letter aggregates as anagrams. They establish that anagram solution time is a function of anagram letter order, solution word frequency (Mayzner & Tresselt, 1958), and transition probability (Mayzner & Tresselt, 1959). By contrasting the solution times of anagrams of high transition probability total (from summed digram frequencies of sequential letter pairs) with those of low transition probability total a positive relationship is established between the two variables. In other words, the higher the transition probability total of a letter aggregate the more difficult it is to alter the letter arrangement and the longer the solution time. Although Mayzner and Tresselt (1959) do not directly infer from these results that the difficulty of altering letter arrangements in words, in contrast to nonsense letter arrangements, is a function of transition probability they do conclude that well patterned letter groups (which they illustrate with words) are more difficult to reorganize than loosely patterned letter groups (which they illustrate with nonsense arrangements) if patterning is defined in terms of transition probabilities.

In the foregoing experiments two variables are considered and to some extent confounded. Words and nonsense letter arrangements used as anagrams differ both in meaning and organization (or patterning). These variables require independent treatment since it is not at all difficult to have nonsense letter arrangements which are more highly patterned than words (if patterning is taken as the equivalent of transition probability total, as Mayzner and Tresselt propose). In the Mayzner and Tresselt studies only nonsense letter arrangements of greater and lesser patterning (transition probability totals) are employed. It is dubious therefore whether the transition probability hypothesis can be applied as an explanation for solution time differences between word and nonsense anagrams, without a test in which words and nonsense arrangements are employed.

The present experiment offers a test of the hypothesis that solution time differences between highly and loosely patterned anagram letter arrangements are a function of transition probability. This is achieved by utilizing both five-letter words and nonsense five-letter aggregates as anagrams, keeping transition probability, letter order, and solution-word frequency constant.

---

[1] Presented at the 1961 meetings of the Eastern Psychological Association in Philadelphia. We express our appreciation to Max Weiner for his assistance in making *S*s available to us. Part of this study was an honors project by Rheba Horn.

METHOD

Subjects.—Sixty students, primarily in their junior year, at Brooklyn College served as *S*s; 30 were randomly assigned to each of two groups which differed in the presentation order of the materials.

Materials.—The stimulus materials consisted of 10 five-letter nonsense aggregates (e.g., OBAVE) typed in capital letters on plain 3 × 5 in. index cards, and another set of 10 five-letter words (e.g., FROTH) typed in identical fashion on index cards. Both words and nonsense letter aggregates are used as anagrams in the experiments.

Experimental design.—One group was first given the 10 nonsense letter anagrams and then the 10 word anagrams; the other group the reverse. The presentation order of individual anagrams was rotated so that every anagram arrangement occurred in each serial position and in a different combination with its counterpart in the other anagram group (i.e., 1,1; 2,3; 3,5; . . . 10,10).

The solution words for both nonsense and word anagrams were selected from the Thorndike-Lorge (1944) word list and equated for frequency. The mean frequencies of nonsense and word anagram solution words are 74.4 and 74.3, respectively.

The traditional anagram problem was modified so that only *two* letters of the nonsense or word anagram were to be interchanged to produce the solution word (keeping the remaining letters in order). The "letter switching order" of the nonsense and word anagram sets was matched to control letter order alteration (i.e., there is one 1–3 switch in each group, one 1–4, one 1–5, etc). This is necessary since some letter orders are more difficult than others (Hunter, 1959); limiting the altering task to two letters controls the unscrambling necessary to solve the anagram.

The anagram sets were also equated for transition probability. The nonsense anagrams were arranged to yield mean transition probability totals equal to the mean totals for the word anagrams. Transition probability totals are derived from summed digram frequencies of sequential letter pairs (Pratt, 1942). For example, the total for the anagram OBAVE is the total of OB, BA, AV, and VE = 4 + 10 + 10 + 29 = 53. The mean transition probability total for nonsense anagrams is 153.8; for word anagrams, 151.7 (the difference is not significant).

The use of transition probability totals from summed digram frequencies is based upon *S*'s supposed knowledge (of probabilities of occurrence) of pairs of letters in his native language. Since *S* may be equally affected by his knowledge of three-letter, four-letter, and five-letter groupings and single letter expectancies in his attempts at solution it suggests that the total transition probability of an anagram should be a composite of digram, trigram, quadragram, and pentagram frequencies, as well as initial and terminal letter frequencies. Pentagram and quadragram frequency tables were not available to us, although trigram frequency tables were (Pratt, 1942). We have computed the mean total of summed trigram frequencies for word and nonsense anagrams. The mean for word anagrams is 204.2 ($SD = 392.2$); for nonsense anagrams, 164.8 ($SD = 355.8$). The difference between these means is not significant. In addition the difference of mean initial letter frequencies is not significant (word mean = 63.0, $SD = 63.1$; nonsense mean = 67.4, $SD = 58.5$), nor the difference between terminal frequency means (words mean = 122.6, $SD = 72.5$; nonsense mean = 72.7, $SD = 81.7$).

Procedure.—The *S*s were told they were to make meaningful words from the nonsense and word anagrams by interchanging two of the letters and keeping the remaining three in the same location. Only one correct solution is possible for each anagram. One example (i.e., CAUSE-SAUCE, ERTEN-ENTER) and one sample problem (i.e., SPOON-SNOOP, FALSH-FLASH) was given before each anagram list. The *S*s were also told there would be 10 anagrams in each list with a maximum 2-min. solution time per anagram, after which the solution word would be given before continuing to the next anagram. Solutions were given by *S* verbally with no aids permitted. Solution times were recorded with a stop watch. Testing was individual and the experiment proper was followed by asking *S*s how they obtained the solutions.

RESULTS AND DISCUSSION

Medians of *S*'s solution times for each anagram set and median group ($N = 30$) solution times per anagram were computed. Table 1 gives the major results expressed as the means and *SD*s of individual medians. Within this design it is possible to evaluate the effects upon solution time of the word-nonsense letter arrangement differences, as well as possible practice

TABLE 1

SOLUTION TIMES (SEC.) FOR NONSENSE ANAGRAM AND WORD ANAGRAM TASKS

Presentation Order	Word Anagrams			Nonsense Anagrams			Total		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
First	30	18.23	21.45	30	8.37	11.19	60	13.30	17.68
Second	30	15.17	12.01	30	10.46	12.75	60	12.81	12.41
Total	60	16.70	17.33	60	9.42	11.83			

Note.—Means and *SD*s are based upon medians of *S*'s solution times for 10 nonsense anagrams and 10 word anagrams.

effects. The word-nonsense arrangement effect was tested with the Friedman two-way analysis of variance by ranks (Siegel, 1956). Word anagram solution time was significantly slower than nonsense anagram solution time ($P < .001$). (First and second order presentations were combined, on the basis of the analysis which follows.)

The effects of practice were assessed separately for the word and nonsense letter arrangement groups. A Mann-Whitney *U* test was applied to the first and second order solution times within each task. The values indicated are not significantly different ($P > .10$).

To assess further the relationship between transition probability and solution time (since transition probability total is already controlled) Kendall τ's were computed between nonsense anagram solution times and their associated transition probability totals. The same was done for the word anagram solution times. The τ's for nonsense anagrams, first and second order presentations, were $-.06$ and $-.16$, respectively. For words, these τ's were .11 and .08. These correlations are not significant. On the other hand, τ between solution times for the first and second presentation of word anagrams was .62, and the comparable τ for nonsense anagrams was .63 ($P < .05$).

The anagram problem solver has two kinds of data to take into account: the anagram and the solution words he generates. If he requires knowledge of the transition probabilities of the letter pairs in his language to treat adequately the letter pairs of the anagrams (Hunter, 1959) the same should apply to the solution words. This is evident from those *S*s who put correct combinations together but do not recognize them as words. (We might be tempted to say that the process of solving anagrams is one of testing varieties of possible letter combinations and then recognizing the combinations as words, were it not that many *S*s produce solution words and *then* check the correctness of the letter combinations.) We tested to be certain the transition probability totals for *solution* words were not different, even though word frequencies were equated. The difference between averaged transition probability totals (from summed digram frequencies) for solution words is not significant (word mean = 145.1, *SD* = 63.9; nonsense mean = 151.4, *SD* = 75.3).

The *S*s divide into those who solve nonsense anagrams faster and a smaller group who solve word anagrams faster. To test whether these represent stable *S* capacities, we first hypothesized that the *S* trait represented here is the ability to deal with embedded figures (viz., letters embedded in words). Such an *S* capacity might be related to an ability to deal with geometric figures embedded in complex designs as represented by the Gottschaldt figures.[2] No such relationship, however, was found to exist.

As Mayzner and Tresselt (1958) and Hunter (1959) show, by different methods, the production of solutions per unit time comes very quickly at first and then tapers off. By applying the method used by Hunter, which indicates the average rate of correct solution production, to a 10-sec. interval, we find the same solution production pattern except that the production rate for word anagrams is slower than for nonsense anagrams. Examination of *S*s' retrospective reports suggests why this may be so. Some *S*s comment that word anagrams often bind them whereas nonsense anagrams do not. These *S*s perseverate on the word anagrams to such an extent that problem solution is interfered with. Whether perseveration is due to attention to the meaning associated with the word or to its sound is not clear.

Word anagrams are more difficult to alter than nonsense anagrams. This supports the findings of Mayzner and Tresselt (1958) and Devnich (1937). Since transition probability was controlled the Mayzner and Tresselt explanation of the difference may be rejected. The low correlation between solution time of individual anagrams and their transition probability totals further supports such rejection. Although solution time differences due to patterning may be accounted for by transition probability it does not follow that it will similarly account for differences in meaning.

Although it has not been possible to compute composite transition probability values for the anagrams, we may conclude that anagram solution time differences between words and nonsense arrangements are not due to averaged digram, trigram, or initial and terminal letter frequencies.

Two possibilities remain to account for the anagram solving differences. The first is the perseverative effect generated by word anagrams and the second, organization or patterning effects independent of transition probabilities of successive letter arrangements.

SUMMARY

An experiment tested the effect of transition probability of successive letter sequences upon the solution time of word and nonsense anagrams. Transition probability totals (from summed digram frequencies of sequential letter pairs), solution word frequencies, and letter switching orders were equated for word and nonsense anagrams.

Solution time is significantly longer for word anagrams than nonsense anagrams. This difference is not accounted for by transition probability totals from summed trigram frequencies, initial and terminal letter frequencies, or transition probability totals of solution words.

Word perseveration effect is offered as a possible explanation of the solution time difference as well as organization effects not accounted for by transition probabilities of successive letter arrangements.

REFERENCES

DEVNICH, G. E. Words as "Gestalten." *J. exp. Psychol.*, 1937, **20**, 297–300.

HOLLINGWORTH, H. L. The conditions of verbal configuration. *J. exp. Psychol.*, 1935, **18**, 299–306.

---

[2] We therefore administered the Thurstone Concealed Figures Test (Thurstone, 1944, 1951) on a group basis to 40 of the 60 *S*s still available to us. The mean correct solutions were 81.9 for nonsense anagrams ($SD = 15.5$) and 84.2 for words ($SD = 17.2$) ($P > .05$). We are indebted to Aaron Carton for suggesting this possibility to us.

HOLLINGWORTH, H. L. Verbal Gestalt experiments with children. *J. exp. Psychol.*, 1938, **23**, 90–95.

HUNTER, I. M. L. The solving of five-letter anagram problems. *Brit. J. Psychol.*, 1959, **50**, 193–206.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56**, 376–379.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47**, 117–125.

NISSENSON, M., & SARGENT, S. S. Words as configurations. *J. exp. Psychol.*, 1941, **28**, 85–89.

PRATT, F. *Secret and urgent.* New York: Blue Ribbon, 1942.

SARGENT, S. S. Thinking processes at various levels of difficulty. *Arch. Psychol., N. Y.*, 1940, No. 249.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Bureau of Publications, 1944.

THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. Chicago Press, 1944.

THURSTONE, L. L. An analysis of mechanical aptitude. *Psychometr. Lab. Bull.*, No. 62. Chicago: Univer. Chicago Press, 1951.

(Received April 19, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 5, 519–520

SUPPLEMENTARY REPORT: A NONVERBAL MEASURE OF EXTINCTION IN SKILL AND CHANCE SITUATIONS[1]

KENNETH B. HOLDEN[2] AND JULIAN B. ROTTER

Ohio State University

James and Rotter (1958) have shown that partial reinforcement does not lead to more resistance to extinction than 100% reinforcement if *S* feels that reinforcements are produced as a consequence of his own skill. The James and Rotter study used verbal expectancies as a measure of learning and extinction. The present study is an attempt to expand the generality of the Internal-External (I-E) variable (Rotter, Seeman, & Liverant, in press) to include nonverbal behavior and therefore be more analogous to the classical extinction studies. This study is essentially a replication of the 50% reinforcement groups of the James-Rotter (1958) study with a betting technique as a behavioral measure of extinction rather than stated verbal expectancies. In addition, a 50% reinforced group receiving neither skill nor chance instructions, but rather instructions neutral on the I-E variable was included.

Method.—The method, except for the betting technique as a measure of extinction, was identical to that used by James and Rotter (1958). The apparatus, stimulus material, and other details of the experiment with the exception of the statement of verbal expectancies were duplicated, with the verbal expectancies being replaced by a money reward procedure. James and Rotter found that the majority of the *S*s responded differentially to skill and chance instructions in a simple card-guessing game, although there were wide individual differences.

The *S*s were seated about 5 ft. in front of the screen containing the shutter of the tachistoscope. Beside the screen was a pile of approximately 400 nickels with several $2.00 rolls of nickels alongside. The *S*s were given $2.00 from this money at the start of the experimental procedure. A white 3 × 5 in. card was taped to the table. Before any cards had been shown, *E* placed a nickel from his pile on the card and *S* was requested to place one from his $2.00 on the card. After a set of seven cards, if *S* had been successful (i.e., had been allowed to correctly anticipate at least five out of seven cards), *E* added a nickel to the card and *S* removed one. If *S* had not been allowed to be successful *E* removed a nickel and *S* had to add one from his original $2.00 if he wished to continue playing. This procedure continued until *S* extinguished or ran out of money.

The James and Rotter (1958) instructions for the skill and chance groups were applied in this study, except for the difference in response measure used. The "ambiguous" group received all the instructions that were *common* to the skill and chance groups except for the statement in the first paragraph of the instructions that "There is some evidence that some people are considerably skilled at this."

The *S*s were 38 men and 38 women, aged 27 yr. or less, from elementary psychology classes during the summer quarter. They were assigned to three groups in a random manner, with the restriction that there would be an equal number of men and women in each group.

Results.—The means and *SD*s for trials to extinction are given in Table 1. The measure was the number of trials (sets of seven card choices) taken by *S*s to extinguish after an initial set of 10 learning trials. A maximum of 40 extinction trials was possible, with 2 of the 76 *S*s failing to extinguish. A nonparametric analysis was used since the distributions did not meet the requirements for *t* tests.

The *SD*s are rather large, especially in the chance and ambiguous conditions, indicating the possibility that the instructions were not as effective for some *S*s as they were for others and that some *S*s operated according to their own hypotheses rather than to the instructions. An *F* test of the ratio of the variances between skill and chance conditions was 1.69, short of the 1.98 necessary for significance at the .05 level. The Mann-Whitney *U* for the difference between the skill and chance conditions gave a *z* of 2.78 (two-tailed $P = .005$). Additionally, the *z* for the difference between the chance and ambiguous conditions was 2.36 ($P = .018$). The *z* of .19 between the skill and ambiguous conditions was, of course, not significant.

It can be seen that, while men tend in the

---

[1] This research was aided in part by the use of funds and data collected for the United States Air Force Contract No. 49(638)-741 monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

[2] Now at San Fernando Valley State College, Northridge, California.

TABLE 1

MEANS AND *SD*s OF NUMBERS OF TRIALS TO EXTINCTION

Conditions	Men			Women			All *Ss*		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
50% Skill	13	13.6	5.1	13	10.7	7.1	26	12.4	6.4
50% Chance	13	18.7	10.7	13	23.2	10.0	26	20.9	10.6
50% Ambiguous	12	17.2	11.1	12	11.2	7.9	24	14.1	10.1

same direction, much of the experimental effect appears to stem from women in the population sampled. The Mann-Whitney *U*s for women were 26 ($P = .002$) and 20 ($P = .002$), respectively, for the differences between the skill and chance and the chance and ambiguous conditions. The corresponding *U*s of 70 and 51 for the men were not significant at the .05 level. It will be noted that in the ambiguous condition the men responded as they did under the chance condition and the women responded as they did under the skill condition.

This study extends the generality of the James and Rotter (1958) findings by showing that the same effects of skill and chance instructions are obtained when a nonverbal response is used to index experimental extinction. There appears to be a sex difference in the effects of the instructions, with the ambiguous instructions giving the same effects as chance instructions with men but giving the same effects as skill instructions with women.

REFERENCES

JAMES, W. H., & ROTTER, J. B. Partial and 100% reinforcement under chance and skill conditions. *J. exp. Psychol.*, 1958, 55, 397–403.

ROTTER, J. B., SEEMAN, M., & LIVERANT, S. Internal versus external control of reinforcements: A major variable in behavior theory. In D. Williams (Ed.), *Decisions, values, and groups*. New York: Pergamon, in press.

(Received December 8, 1960)

ERRATUM

In the article "The Effect of Stimulus Similarity on the Acquisition and Extinction of a Conditioned Response" by D. P. Hunt (*J. exp. Psychol.*, 1962, 63, 278–283) the line in the fourth paragraph, third sentence (p. 278) should read as follows: "By assuming equal terminal levels of I_n. . . ."

Journal of
Experimental Psychology

VOL. 63, NO. 6 JUNE 1962

MINIMAL PAIRED-ASSOCIATE LEARNING[1]

LLOYD R. PETERSON AND MARGARET JEAN PETERSON

Indiana University

Paired-associate learning can be conceptualized as complex discrimination learning. The simplest analysis should be possible where the number of pairs is reduced to a minimum of two, in which case the parallel to simple discrimination learning is apparent. Presumably paired-associate learning in this limiting case has not been investigated because it is too easy, and therefore a proper learning curve could not be obtained. However, it has been shown that an even simpler phenomenon, recall of single items, can be studied effectively by using retention over short intervals as the measure of response strength (Peterson & Peterson, 1959; Peterson, Peterson, & Miller, 1961). The extension of this technique to paired associates requires the addition of a cue for the response as well as a second response and its cue. The consideration that a specific cue is provided for the response might be expected to improve retention, but with the introduction of a second response the difficulty of the task should be increased. The series of experiments to be reported investigated the interference effects from paired associates presented either just before or just after the pair to be tested.

GENERAL METHOD

Subjects.—The *S*s were students from introductory psychology courses at Indiana University. Participation in experiments was a course requirement.

Apparatus.—A memory drum was mounted in a 3 × 4 ft. black screen. A sliding aperture permitted exposure of seven columns each containing 28 spaces. The drum motor was controlled by two decade interval timers.

Test materials.—The materials were pairs of English words having frequencies of occurrence of at least 10 per million (Thorndike & Lorge, 1944). The first member of a pair was a three-letter word typed in red capital letters, and the second was a four-letter word typed in black lower case letters. The words were paired randomly with the restriction that pairings in which the words were obviously related were eliminated. Pairs tested for recall appeared equally often in each condition of an experiment over a group of *S*s. A given pair was presented only once to an individual *S*.

Procedure.—The *S*s were instructed to look at the stimulus words and to pronounce aloud the black four-letter response words in Exp. I. In Exp. II through IV *S*s were told to pronounce aloud both stimulus and response words. The *S*s were instructed that each

---

[1] This research was supported by Grant G12917 from the National Science Foundation to Indiana University for which the senior author is principal investigator.

TABLE 1

Analysis of Variance of Mean Proportions Correct in Exp. I

Source	*df*	*MS*	*F*
Presentation (P)	2	3.383	26.02**
Retention (R)	2	.237	8.46**
*S*s	26	.110	
P × R	4	.014	.48
P × *S*s	52	.130	4.48**
R × *S*s	52	.028	.97
P × R × *S*s	104	.029	

** $P < .01$.

stimulus word was paired with a response word and that the pair was to be learned. Words were exposed one at a time for about 1 sec. each. A stimulus word appeared followed by a response. Depending on the condition one or two other pairs might follow. A randomly selected three-digit number then appeared, following which the drum stopped on a blank space. The *S*s counted backward by 2, 3, or 4 as rapidly as possible when the number appeared and continued until the signal for recall. Instructions for a change in counting were given during 1-min. rest periods coincident with a change of the tape. The signal for recall was a red light in Exp. I and a bell in Exp. II through IV. Accompanying this signal was an exposure of one of the stimulus words, and *S* was instructed to recall the response word that had been previously paired with this stimulus. The retention interval was measured from removal of the response to be tested from the aperture to appearance of the cue for recall. Fifteen seconds after the signal for recall in Exp. I, 12 sec. in Exp. II through IV, a new cycle was begun by the flashing of a green light as a blank space was exposed. There followed a new stimulus term. Practice tests were given following the instructions, in which *S*s were corrected for procedural errors and told to count as fast as possible if they hesitated.

EXPERIMENT I

Method

Each of 27 *S*s was tested six times in every condition of a 3 × 3 factorial design consisting of three conditions of presentation tested at retention intervals of 4, 8, and 16 sec. In Cond. S (Single) only one pair of words was presented. In Cond. DA (Double A) two pairs were presented and the first was tested. In Cond. DB (Double B) two pairs were presented and the second was tested. Each condition appeared once per block of nine tests in a random position. The *S*s pronounced the response but not the stimulus. Six practice tests were given.

Results and Discussion

The analysis of variance summarized in Table 1 indicates that both retention interval and presentation condition influenced correct recalls. Recall in the two-pair conditions (Table 2) was markedly lower than that in the one-pair conditions, as would be predicted from considerations of interference. The superiority of DA over DB is a complicating circumstance, although the Presentation × *S*s interaction appears to indicate an absence of the majority trend with some of the *S*s.

Errors are classified as to source in Table 2. The term "alternate" refers

TABLE 2

Mean Proportions Recalled in Exp. I

Measure	Cond. S Intervals			Cond. DA Intervals			Cond. DB Intervals		
	4	8	16	4	8	16	4	8	16
Correct	.94	.89	.84	.67	.63	.60	.57	.46	.43
Alternates				.27	.22	.26	.35	.39	.36
Stimuli	.01	.00	.01	.00	.01	.01	.02	.02	.02
Previous	.01	.04	.04	.01	.06	.05	.03	.04	.06
Extraneous	.03	.00	.00	.01	.02	.01	.01	.00	.03
Omissions	.01	.07	.11	.04	.06	.07	.02	.09	.10

to the nontested response member of the current cycle, and "stimuli" are those from that same cycle. "Previous" errors include stimuli and responses from earlier cycles. "Extraneous" errors include any overt response not previously used in the experiment. It is apparent that recall of the alternate response accounts for the bulk of the errors of commission. The greater incidence of alternate responses presented first as compared to those presented second is reversed from what would be predicted on the basis of recency, for the time since presentation of this competing response is longer in Cond. DB than in Cond. DA.

An explanation of the primacy effect might be that *S*s were required to read aloud only the responses, and thus they may have failed to reliably observe the second stimulus. A test of this hypothesis was made by requiring *S*s to read aloud both stimuli and responses. In addition, the number of pairings was varied to ascertain if repetition might influence the primacy effect.

TABLE 3

ANALYSIS OF VARIANCE OF MEAN PROPORTIONS CORRECT IN EXP. II

Source	*df*	*MS*	*F*
Repetition (R)	1	2.083	40.84**
Order (O)	1	.631	7.09*
Half (H)	1	.148	2.51
Ss	23	.163	
R × O	1	.020	
R × H	1	.005	
R × *Ss*	23	.051	1.34
O × H	1	.020	
O × *Ss*	23	.089	2.34**
H × *Ss*	23	.059	1.55
R × O × H	1	.099	2.61
Error (res.)	92	.038	

* $P < .05$.
** $P < .01$.

EXPERIMENT II

Method

A 2 × 2 within-*S*s design was used with 24 *S*s. A set of two pairs was presented either once or twice, and either the first or second pair was tested. In the repetition conditions three orders of presentation occurred equally often for each *S*: AABB, ABAB, and ABBA. Each of the four main conditions appeared equally often in a random position in each block of 12 tests. Each condition was tested 12 times following 4 practice tests. The retention interval was 8 sec., and the intercycle interval was 12 sec.

Results and Discussion

The data were analyzed in such a way that first and second halves of the session could be compared, but the analysis of variance (Table 3) did not show this to be a significant variable. The primacy effect persisted with repetition, as can be seen in Table 4,

TABLE 4

MEAN PROPORTIONS RECALLED IN EXP. II

Measure	First Half				Second Half			
	One Rep.		Two Rep.		One Rep.		Two Rep.	
	A	B	A	B	A	B	A	B
Correct	.40	.38	.69	.53	.53	.38	.71	.60
Alternates	.14	.12	.12	.20	.08	.17	.13	.19
Stimuli	.06	.05	.06	.04	.05	.08	.06	.04
Previous	.12	.08	.03	.05	.06	.07	.02	.01
Extraneous	.02	.04	.02	.01	.03	.02	.04	.03
Omissions	.26	.33	.08	.17	.25	.28	.04	.13

the repetition increasing recall for both first and second pairs. The *S*s rarely made errors in reading, and no errors involved the second pair, so there can be no question but that *S*s observed the second response and its cue. The primacy effect persisted with all sequences of presentation in the repetition conditions. Proportions correct when the first pair was tested were .71, .68, and .70 for the AABB, ABAB, and ABBA sequences, respectively. The proportions in testing of the second pair were .55, .57, and .58, in the same order.

Recall in the one-presentation conditions was lower than in Exp. I at the 8-sec. retention interval. This decrease may be related to changes in error frequencies seen by comparing Tables 2 and 4. Occurrences of alternate responses were sharply reduced in Exp. II, while stimulus errors increased somewhat. These changes may be related to the fact that *S* read the stimuli aloud in Exp. II. Frequency of words from previous cycles increased somewhat in Exp. II, probably reflecting vocalization of stimuli, repeated presentations, and a decrease in time between cycles. Errors from previous cycles decreased slightly in the second half of the session, and in either session accounted for only a small proportion of the errors. The largest change from Exp. I to Exp. II was an increase in the number of tests in which no response was given, indicating that the overt errors provide only a partial analysis of the forgetting.

TABLE 5

ANALYSIS OF VARIANCE OF MEAN PROPORTIONS CORRECT IN EXP. III

Source	*df*	*MS*	*F*
Order (O)	2	.227	8.74***
Half (H)	1	.181	6.46*
*S*s	35	.154	7.33
O × H	2	.016	.76
O × *S*s	70	.026	1.24
H × *S*s	35	.028	1.33
O × H × *S*s	70	.021	

* $P < .05$.
*** $P < .001$.

TABLE 6

MEAN PROPORTIONS RECALLED IN EXP. III

Measure	First Half			Second Half		
	A	B	C	A	B	C
Correct	.28	.24	.18	.36	.26	.25
Alternates	.13	.18	.18	.18	.16	.15
Stimuli	.09	.07	.05	.07	.05	.08
Previous	.08	.09	.09	.05	.06	.08
Extraneous	.02	.07	.05	.02	.03	.03
Omissions	.40	.35	.45	.32	.44	.41

EXPERIMENT III

The persistence of the primacy effect in Exp. II led to an investigation of recall with three pairs of associates. It was considered possible that the middle position might prove to be inferior in recall to both first and last positions as in serial list learning and immediate memory experiments (Broadbent, 1958).

Method

Three pairs were presented once, and after a retention interval of 12 sec., the bell sounded at the same time that one of the three stimuli appeared to cue recall. Each condition was tested 16 times with each of 36 *S*s. The sequence of conditions was randomized within each of four blocks of 12 tests. The interval from recall cue to next cycle was 12 sec. Four practice cycles were given.

Results and Discussion

The analysis of variance summarized in Table 5 shows that again order in presentation was a significant variable. However, no bowing of the position curve was found (Table 6). There was a straight primacy effect.

There was a significant increase in recall in the second half of the session, but this did not interact with presentation order.

The alternate errors tabulated in Table 5 include both of the other response members within the cycle being tested, since no differences between them appeared reliable. Again, errors from the current cycle predominated, and again there was a slight decrease in errors from previous cycles during the second part of the session.

EXPERIMENT IV

The previous experiments showed a decrement in recall related to the number of pairs within a cycle. Response interference seemed a likely explanatory factor, but it did not account for the persistent serial position effect. A factor influencing recall which might produce a primacy effect is stimulus generalization decrement. If the stimulus situation for the response presented last were to be considered to extend in time to include the preceding pair of words, then the stimulus at recall differs in that the two preceding words are now absent. Stimulus conditions at recall for the pair presented first are more similar to those of presentation than for the second pair, since no words precede its cue in either presentation or recall. Thus, superiority of recall would be predicted for the first pair by reason of generalization decrement. An experiment was designed which increased the number of cues at recall by presenting the alternate pair as part of the stimulus situation.

Method

Fifty-six *S*s were assigned in order of appearance to a randomized list partitioning them into two groups. Group S (Silent) was instructed to look at the words which appeared at recall but not to speak them aloud. Group V (Vocal) was instructed to speak each word aloud. Both groups read the words in the original presentation aloud. Two pairs of words were presented in each cycle. Each *S* served in all conditions of a 2 × 2 factorial design, in which one factor was the serial position during presentation, and the other factor was the presence or absence of members of the other pair as cues for recall. For half of the 48 tests three of the previously presented words appeared one by one in the drum aperture in the order: stimulus, response, stimulus. When the drum stopped at a blank, *S* attempted to supply the correct response to the last stimulus. On the remaining tests, when the bell rang the drum showed two successive blanks, the single stimulus for the response tested appeared, and the drum stopped on a blank. The interval to beginning of the next cycle was 12 sec. Four practice cycles were given. The retention interval was 12 sec.

Results and Discussion

The analysis of variance summarized in Table 7 shows that there was, as predicted, a cue effect in recall which interacted with order in

TABLE 7

ANALYSIS OF VARIANCE OF MEAN PROPORTIONS CORRECT IN EXP. IV

Source	*df*	*MS*	*F*
Between			
Vocalization (V)	1	.010	.05
Error (b)	55	11.319	
Within			
Order (O)	1	1.856	35.02***
Cues (C)	1	.126	3.15
Half (H)	1	1.568	39.20***
O × C	1	.536	15.76***
C × H	1	.003	.09
O × H	1	.033	.97
O × V	1	.007	.21
C × V	1	.007	.21
H × V	1	.022	.65
O × C × H	1	.000	
O × C × V	1	.052	1.53
O × H × V	1	.052	1.53
C × H × V	1	.002	.06
O × C × H × V	1	.008	.24
Error (O)	55	.053	
Error (C)	55	.040	
Error (H)	55	.040	
Error (res.)	212	.034	

*** $P < .001$.

TABLE 8

MEAN PROPORTIONS RECALLED IN EXP. IV

Measure	First Half				Second Half			
	Pair A		Pair B		Pair A		Pair B	
	One Cue	Three Cues	One Cue	Three Cues	One Cue	Three Cues	One Cue	Three Cues
Group S (Silent)								
Correct	.49	.51	.28	.37	.58	.55	.42	.51
Alternates	.08	.05	.14	.07	.15	.07	.18	.03
Stimuli	.05	.01	.06	.04	.04	.02	.07	.03
Previous	.08	.05	.03	.09	.06	.05	.06	.09
Extraneous	.04	.05	.05	.06	.02	.04	.01	.06
Omissions	.26	.33	.44	.37	.15	.27	.26	.28
Group V (Vocal)								
Correct	.48	.41	.27	.39	.61	.55	.40	.51
Alternates	.09	.00	.16	.03	.10	.01	.18	.04
Stimuli	.05	.01	.06	.01	.03	.00	.10	.00
Previous	.08	.14	.07	.10	.06	.07	.05	.06
Extraneous	.02	.04	.02	.05	.02	.02	.03	.04
Omissions	.28	.40	.42	.42	.18	.35	.24	.35

presentation. Table 8 indicates that the additional cues increased recall for the pair presented second, but had no consistent influence on pairs presented first. Inspection of the error analysis indicates that occurrence of alternate responses and stimuli from the current cycle decreased under the additional cues for both positions in presentation. The increase in omissions for the A conditions with additional cues suggests that under these conditions *S*s were able to refrain from speaking certain errors, but were not able to replace these errors with the correct response in the 12 sec. before the next cycle began. Although the hypothesis concerning the effect on serial order of stimulus generalization was supported by the data, the persisting small primacy effect suggests that other factors are also present.

The series of experiments taken together support an interference interpretation of the forgetting which occurs over short intervals of time. Little forgetting occurred with one response in a cycle, and hence proactive interference from items presented in earlier cycles does not seem of major importance. Such interference from earlier items as was evidenced by the analysis of errors, did not increase from the first half to the second half of the session. On the other hand, pairs presented either immediately before or immediately after presentation of the tested pair did markedly decrease retention. The data suggest that recency is a major variable in the relationship between previous learning and new learning.

SUMMARY

The short-term retention of minimal numbers of paired associates was studied in four experiments. Pairs of words were presented in a memory drum and tested for recall by presenting a stimulus after some interval no longer than 16 sec. during which *S*s

counted backward. Recall decreased as number of pairs increased, but a primacy effect complicated an interpretation in terms of competing responses. This primacy effect was present with either one or two presentations of a pair. When the stimulus situation at recall was extended to include the alternate pair of words, the primacy effect was reduced. Errors at recall were largely words from the set of pairs just presented.

REFERENCES

BROADBENT, D. E. *Perception and communication*, New York: Pergamon, 1958.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

PETERSON, L. R., PETERSON, M. J., & MILLER, A. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received May 1, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 528–533

STUDIES IN CONCEPT ATTAINMENT: III. EFFECT OF INSTRUCTIONS AT TWO LEVELS OF INTELLIGENCE [1]

SONIA F. OSLER

Goucher College

AND SANDRA RAYNES WEISS

Department of Pediatrics, Johns Hopkins University

In a recent study on the effect of intelligence on concept attainment using instructions which were nonspecific as to the nature of the task, Osler and Fivel (1961) reported a significant difference in errors between groups of average and superior IQ. An important question that has yet to be answered concerns the pertinent attribute of high intelligence which accounts for the superiority of performance. One possibility is that superior *S*s supplement *E*'s instructions with their own, directing them to search for consistencies in the reinforced stimuli; whereas the less intelligent *S*s work along without self-instructions until the reinforcement contingencies of the experiment strengthen the response to the concept exemplars. Perhaps a simpler statement of this hypothesis might be that under the conditions of nonspecific instructions the task presented to *S* has two components, namely, problem *finding* as well as problem *solving*; and that the superior performance associated with high intelligence may be the result of greater proficiency at either or both of these components. On the other hand, when instructions are specific, defining the problem which *S* is expected to solve, the problem finding component of the task is eliminated or reduced to a minimum.

It was the aim of the present investigation to test these ideas experimentally. Two experiments were performed. Experiment I was a replication of the Osler and Fivel (1961) experiment. This was undertaken for the double purpose of confirming the earlier data and controlling for the possibility that automated equipment presently being used for the conduct of the experiment may have affected the results obtained. In Exp. II the general in-

---

[1] This investigation was supported by a grant from the Psychopharmacology Service Center, National Institute of Mental Health, made to Leon Eisenberg, Johns Hopkins University, and the senior author (Grant MY-2583).

The writers wish to express their gratitude to Mary A. Adams, Assistant Superintendent, Elementary Education; Vernon S. Vavrina, Assistant Superintendent, Secondary, Vocational, and Adult Education; and to Arthur Lichtenstein, Director of Special Services for Pupils, Baltimore City Department of Education; and to B. Melvin Cole, Director, Elementary Education; Norris A. King, Director, Secondary Education; and Charles J. Leiman, Supervisor, Psychological Services; and to Loyal W. Joos, Director, Educational Research and Planning, Baltimore County Department of Education; and to the principals and their staffs who cooperated in the conduct of the study.

The authors also wish to thank Roy Lachman, Sandra L. Shapiro, and Grace E. Trautman for testing *S*s and for many suggestions concerning the conduct of the study. To Margaret P. Martin, Associate Professor of Biostatistics, Johns Hopkins School of Hygiene and Public Health, go our special thanks for her invaluable assistance with the statistical problems encountered in this investigation.

structions were replaced by more specific ones, defining the problem for *S* and directing him to search for hypotheses.

METHOD

As Exp. I and II differed only in the instructions used, the method for both will be described together, and wherever necessary information pertinent to each experiment will be provided separately.

Subjects.—The *S*s consisted of 180 school children attending elementary and junior high schools of Baltimore City and County, equally divided among the following age groups: 6, 10, and 14 yr. Each *S* was within 6 mo. of the age specified. Half of the *S*s within each age level had a WISC IQ between 90 and 109 (Average Intelligence group), and the other half had a WISC IQ of 110 and above (High Intelligence group). The mean IQs of the two groups in Exp. I were 100.8 and 120.9, and in Exp. II they were 101.2 and 121.7, respectively.

Apparatus.—The apparatus has been described elsewhere (Osler & Powell, 1960). Essentially, it consisted of a film strip projector and a screen to receive the projected images of the stimuli which appeared in pairs. A telegraph key placed under each stimulus released a marble following a correct response, which consisted of pressing the key under the concept exemplar. No reinforcement followed if the key under the negative stimulus was pressed. The *S*s were not paced, but 4 sec. after a response was made the next stimulus was presented.

The concept studied was *bird.* This was one of the three concepts on which data were reported by Osler and Fivel (1961). The stimuli used were the same as in the previous investigation, except that they now appeared on a screen projected from a film strip. The complete set of stimuli consisted of 150 pairs of pictures, one member of each pair representing a positive instance of the concept and the other a negative instance. The pictures of each pair were matched approximately in area, color, and brightness. Reasons for using naturalistic stimuli, which carry with them the hazard of unspecified dimensions, have been discussed elsewhere (Osler & Fivel, 1961). An additional reason prompting their use in this investigation was the previous finding that performance on this concept was a function of intelligence, thus demonstrating the suitability of this problem for the present study. The stimuli had been pretested on a group of children of the same age and intelligence levels as the present *S*s, and were found to be familiar to the type of subject studied.

Procedure.—Each *S* was tested individually in two separate sessions. At the first session *S* was told that he would play a game in which he could win any one of a group of toys displayed before him. He was asked to select the most preferred of the group, and this was then set slightly apart from the others in a position which allowed *S* to view it during the "game." The *S*s in Exp. I received the following instructions:

> This is the game we are going to play. Here are two handles (*E* points to them). If you push the correct handle you will get a marble. Sometimes this will be the correct handle and sometimes this will be the correct handle. If you push the correct handle you will get a marble like this (demonstrated). If you push the wrong handle, like this, no marble will come (demonstrated). If you get a marble you can put it here (here *E* points to a grooved wooden tray with which *S* was provided). Now you try this one and let's see what happens. If you look at these pictures before pushing the handle, you will be able to get many marbles. When you fill the tray with marbles, you can have the (name of the toy).

The stimuli used during the demonstrations were unrelated to the concept.

In Exp. II the *S*s were also informed about a game in which they could win a toy, and then they received the following instructions:

> This is the game we're going to play. Listen carefully and I will tell you how to play the game. Here are two pictures. You pick one of them and push the handle under it, and if you are right, you will get a marble, like this (demonstrated). If you are wrong you will not get a marble, like this (demonstrated). Now you try and push this one. The game is to see how soon you can get a marble every time you choose. If you look at the pictures carefully you will see that there is something in the pictures like an idea that will tell you which one to choose to get a marble every time. Now we will start the real game. What are you going to do first? (if *S* does not say: look at the pictures, *E* supplies the answer). If you look carefully and see which pictures give you the marble you will be able to win marbles every time. Are you ready? Remember, look at the pictures first. (The preceding sentence

TABLE 1

MEAN NUMBER OF ERRORS TO A CRITERION OF 10 CONSECUTIVE CORRECT RESPONSES WITH GENERAL INSTRUCTIONS

Intelligence	Groups						Means by IQ
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Average IQ	71.2	21.0	38.9	34.5	14.3	22.2	41.5
High IQ	57.1	27.8	19.7	17.0	6.5	11.4	27.8
Means by Age	64.2		29.3		10.4		

Note.—$F_{\text{age}} = 37.66$, $df = 2$, $P < .001$; $F_{\text{Intel.}} = 7.14$, $df = 1$, $P < .01$; $F_{\text{Age} \times \text{Intel.}} = <1$.

was repeated together with the following sentence whenever necessary.) Remember, there is an idea that will tell you which picture will give you the marble.

The procedure was continued until *S* made 10 consecutively correct responses, which was the criterion of success. At this point *E* commented on *S*'s good work and asked him how he knew where the marble was each time. If *S* could not state the rule correctly the procedure was continued with questions following each block of 10 consecutively correct responses until *S* gave a correct answer or completed 150 trials. Any *S* who did not achieve criterion performance at that point was considered to have failed. A second session, held within 2 weeks of the first, was devoted to the administration of the WISC.

Experimental design.—In both experiments a 3 × 2 factorial design was used with three ages and two intelligence levels constituting the respective variables. Thus there were six experimental groups with 15 *S*s in each group, a total of 90 *S*s in each experiment.

RESULTS

The results were analyzed in terms of three measures of performance: (*a*) the mean number of errors to criterion, (*b*) the number of *S*s who attained the criterion, and (*c*) the number of *S*s who verbalized the concept at the point of success.

Experiment I.—The mean number of errors per group and the results of an analysis of variance of errors,[2] presented in Table 1, indicate that age and intelligence were associated with reliable differences in performance, thus reproducing the findings reported by Osler and Fivel (1961). Similar results were obtained from an analysis of the data on frequency of *S*s who attained criterion performance and those who were able to verbalize the concept, as shown in Table 2. In summary, then, it turned out that with instructions that did not specify the nature of the task, age and intelligence were associated with improved achievement in concept attainment on each of three measures of performance.

Experiment II.—Table 3 presents the mean number of errors per group and the results of an analysis of variance of errors, respectively, for Exp. II. In this experiment, in which the nature of the task was explicitly defined, the effect of intelligence disappeared; only age was associated

---

[2] Among the 14-yr-old *S*s the conditions necessary for application of an analysis of variance were not entirely satisfied because of skewness of the distribution and the smaller variance of the 14-yr-old High Intelligence group. For this age group the distribution of errors was examined and compared (by χ^2) for average and high intelligence within and between the two experiments. There was no evidence of significant differences among the 14-yr-old *S*s according to intelligence or between the experiments.

TABLE 2

Number of *S*s Who Reached Criterion and Number Who Verbalized the Concept under General Instructions

Intelligence	Groups						All Ages	
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s			
	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	4	1	8	2	13	13	25	16
High IQ	7	4	14	11	15	14	36	29
Total	11	5	22	13	28	27	61	45

Note.—Criterion data: Age $\chi^2 = 22.690$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 6.156$, $df = 1$, $P < .02$. Verbalization data: Age $\chi^2 = 33.068$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 7.512$, $df = 1$, $P < .01$.

with significant differences in the number of errors. Similar results were obtained from an analysis of the other two measures of performance, number of *S*s who achieved criterion performance and number of *S*s who verbalized the concept, as shown in Table 4.

Combined data from the two experiments.—In view of the comparable variances obtained in both experiments the error data were combined and analyzed for a possible Instructions × *S* interaction. The results of this analysis, presented in Table 5, show that while instructions produced no main effect on performance, the Age × Instruction interaction was significant at the .05 level. When the data were subsequently analyzed separately for each age group, it turned out that only for the 6-yr. *S*s did instructions produce a significant difference in performance ($F = 5.92$, $df = 1.59$, $P < .05$) in favor of explicitness. Another finding, rather surprising, and one difficult to explain, was that the bright 10-yr. *S*s were *handicapped* by the explicit instructions. The mean number of errors for these *S*s under the two conditions was 19.7 and 37.0, respectively ($t = 2.57$, $df = 28$, $P < .02$). The oldest group, the 14-yr. *S*s, performed identically under both conditions. However, as is apparent from their scores, these *S*s performed near the maximum possible level, and, therefore, could not

TABLE 3

Mean Number of Errors per Group for Reaching Criterion of 10 Consecutive Correct Responses with Explicit Instructions

Intelligence	Groups						Means by IQ
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Average IQ	46.1	25.5	34.2	26.5	14.1	17.4	31.5
High IQ	46.5	34.0	37.0	31.3	7.7	17.4	30.4
Means by Age	46.3		35.6		10.9		

Note.—$F_{\text{age}} = 13.52$, $df = 2$, $P < .001$; $F_{\text{intel.}} < 1$; $F_{\text{Age} \times \text{Intel.}} < 1$.

TABLE 4

NUMBER OF *S*s WHO REACHED CRITERION AND NUMBER WHO VERBALIZED THE CONCEPT UNDER EXPLICIT INSTRUCTIONS

Intelligence	Groups						All Ages	
	6-Yr. *S*s		10-Yr. *S*s		14-Yr. *S*s			
	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	11	6	12	8	14	12	37	26
High IQ	9	4	9	9	14	14	32	27
Total	20	10	21	17	28	26	69	53

Note.—Criterion data: Age $\chi^2 = 7.081$, $df = 2$, $P < .05$; Intel. $\chi^2 = 1.552$, $df = 1$, *ns*. Verbalization data: Age $\chi^2 = 17.715$, $df = 2$, $P < .001$; Intelligence $\chi^2 = .046$, $df = 1$, *ns*.

be expected to show improvement under any conditions.

The analysis of the combined frequency data of successful *S*s shows that aside from age, reliable differences were produced by instructions in the Average Intelligence but not in the High Intelligence group. The same results were yielded by the data on correct verbalization.

Rank-order correlations were computed between error scores and MAs in each experiment. The rho values obtained were −.68 under general instructions and −.51 under explicit instructions. Both values are significant at the .001 level. Since the MA is heavily determined by the CA, the value of the correlation per se does not throw light upon the role of the IQ alone. However, as *S*s in both experiments were identical in age and intelligence, the difference between the two values of rho may be attributed to variation in experimental conditions. Unfortunately, no statistical test exists for evaluating the difference between the rho values; therefore all that can be said is that the direction of the difference is consistent with the other findings.

TABLE 5

ANALYSIS OF VARIANCE OF ERRORS FOR EXP. I AND II COMBINED

Source	*df*	*MS*	*F*
Instructions (I)	1	616.05	
Age (A)	2	29,819.62	45.05**
Intelligence (IQ)	1	2,456.81	3.71
I × A	2	2,382.52	3.60*
I × IQ	1	1,798.67	2.72
A × IQ	2	7.90	
I × A × IQ	2	407.87	
Within	168	661.89	
Total	179		

* $P < .05$.
** $P < .001$.

DISCUSSION

Under general instructions high intelligence was associated with superior achievement in all three measures of performance; but this superiority was not maintained under explicit instructions. If the assumption that specific instructions eliminate the problem-finding component from the task is correct, then the conclusion to be drawn is that high intelligence gave *S* an advantage in problem definition but not in problem solution in the concept attainment task. It is entirely possible, however, that were the concept task one of greater difficulty, intelligence might also prove to be advantageous in the solution phase of the task. This requires further investigation.

Since age was associated with more effective concept attainment on all three measures of performance and under both types of instructions, it may be concluded that maturation involves growth in problem *solving* ability at least until some point between 10 and 14 yr.

In order to assess changes in problem *finding* ability as a function of age, it is necessary to compare performance at each age level under the two types of instructions. That 6-yr. *S*s improved under the explicit instructions is apparent from the significant Age × Instruction interaction in Table 5. The situation was not clear for the 10-yr. *S*s because they actually deteriorated under explicit instructions. As this finding seemed to be inexplicable on any logical basis and was inconsistent with the results pertaining to the other two age groups, it was decided to replicate this portion of Exp. II. Upon replication it turned out that the average 10-yr. *S*s profited from the explicit instructions, while the superior *S*s were unaffected. As to the 14-yr. *S*s, whose performance was the same under both conditions, they did not have much room for improvement, precluding the possibility of reaching a definite conclusion on the effect of instructions at this age.

It thus appears that at least between 6 and 10 yr. *S*s benefited from explicitness of instructions. It is therefore reasonable to conclude that the more effective performance associated with increasing age is not brought about by improvement in problem finding, but rather by more effective problem solving. This points to a possible interesting difference between mental age and intelligence, the effectiveness of the latter being due to more effective problem finding rather than problem solving.

SUMMARY

A concept attainment task was presented under general and explicit instructions to *S*s 6, 10, and 14 yr. of age, divided into two intelligence groups. The results were analyzed in terms of three measures: errors, number of *S*s who achieved the criterion of success, and number of *S*s who verbalized the concept correctly.

Under nonspecific instructions superior intelligence was associated with more effective concept attainment on three measures of performance, but under explicit instructions *S*s of average intelligence improved while those of superior intelligence remained unchanged. Apparently, superior intelligence gave *S*s an advantage in the problem finding phase of the task, but not in actual problem solution.

Performance improved throughout the age range studied. There was no evidence for improved problem finding ability between ages 6 and 10 yr., and it was difficult to evaluate the data for the 14-yr-olds. It appears, therefore, that problem solving ability, but probably not problem finding ability, increases within the age range studied.

REFERENCES

OSLER, S. F., & FIVEL, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *J. exp. Psychol.*, 1961, **62**, 1–8.

OSLER, S. F., & POWELL, M. G. Apparatus for the study of discrimination and concept formation. *Amer. J. Psychol.*, 1960, **73**, 627–629.

(Early publication received December 18, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 534–537

INHIBITORY POTENTIAL IN ROTARY PURSUIT ACQUISITION BY NORMAL AND DEFECTIVE SUBJECTS[1]

R. WAYNE JONES AND NORMAN R. ELLIS

Louisiana State University *George Peabody College for Teachers*

One systematic approach to the problem of behavioral subnormality is represented by attempts to utilize descriptive or "explanatory" models used in the study of the normal organism. Constructs such as habit strength, motivation, inhibition, generalization, etc., may serve to delimit a class of functions within which individual differences may be most evident. For example, it is possible that the defective's handicap involves mainly an associative mechanism. On the other hand, the principal deficit may be motivational with associative and other components contributing secondarily if at all. The present experiment adopts this general orientation. The concern here is with inhibition.

A previous study (Ellis, Pryer, & Barnett, 1960) showed that normal *S*s gained relatively more during a rest interval than did mental defectives while performing on the rotary pursuit task. The defining operations for this gain were similar to those used by other theorists (Ammons, 1947; Hull, 1952; Kimble, 1949) to describe the dissipation of inhibition. On this basis, it was hypothesized, ad hoc, that the defective built up less inhibition (I_R) than did the normal *S*. Hull's model was used since the general notion of fatigue seemed most germane. From this hypothesis Ellis et al. (1960) predicted that spaced as compared to massed learning should be *relatively* more favorable to normal than to defective *S*s. The present experiment is designed to test the hypothesis that there are differences in the rate of accumulation of inhibition in these *S*s which are independent of performance skill. These differences should appear as interaction effects.

METHOD

Subjects.—The 160 *S*s consisted of 80 institutionalized mental defectives and 80 normal high school students. Equal numbers of males and females occurred in each group. The defectives ranged in IQ (Binet or Otis) from 60 to 75, and in CA from 13 to 28 yr. Those relatively free of gross motor and sensory defects were selected. The IQs of the tenth grade high school students were not available, but only those with a grade average above "C" were used. They ranged in CA from 14 to 16 yr.

Apparatus.—The apparatus was a pursuit rotor (Lafayette Instrument Company, Model No. 203) and an electric stopclock calibrated in .01 sec. The rotor turntable was 10 in. in diameter with a $\frac{3}{4}$ in. round brass target mounted flush in the surface $3\frac{1}{2}$ in. from the center. The turntable rotated clockwise in a horizontal plane at 60 rpm. The on-off periods were controlled automatically by a repeat cycle timer.

Procedure.—The 80 *S*s in each group were assigned equally and without bias to two treatments. Sexes were represented equally in each of the four subgroups. Massed practice was defined by 10-sec. intertrial intervals and distributed practice by 30-sec. intervals. A trial was 20 sec. practice. Each *S* received a block of 30 trials, a 5-min. rest, and a postrest block of 20 trials. The *S*s' behavior was not controlled during the rest except, of course, they were not permitted to engage in any observable type of practice. General encouragement was given throughout practice.

---

[1] This study was supported by Grants M-2229 and M-4433 from the National Institute of Mental Health, Public Health Service, made to the second author. A thesis based on this investigation was submitted by the first author to the graduate faculty, Louisiana State University, in partial fulfillment of the Masters of Arts degree. Thanks are due Margaret W. Pryer, Shirley Wrightsman, and Janice Robinson for statistical computations.

In order to avoid "floor" effects in performance, *S*s failing to score as much as 3 sec. on target for one of the last two trials in the first block were excluded. This criterion eliminated 71 defectives and 4 normal *S*s. The defectives were 16 males and 39 females in the massed group and 5 males and 11 females in the spaced group. The 4 normal *S*s were females and in the massed group.

RESULTS AND DISCUSSION

Figure 1 depicts the performance of the four groups for the pre- and postrest trials. All curves are fairly typical for this task, consisting of linear prerest improvement, a reminiscence effect, and postrest warm-up. Table 1 presents a summary of an analysis of variance of Prerest Trials 11–30.

The A × B interaction shown in Table 1 is critical to the hypothesis of a differential build-up of I_R. The *F* of 3.67 approaches the .05 level (3.92 required with $df = 1/150$). Other results summarized statistically in Table 1 are predictable from previous work in this area. The sex effect will receive attention later.

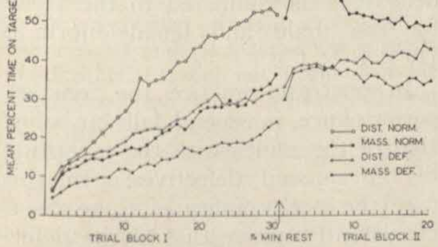


FIG. 1. Prerest and postrest performance curves.

TABLE 1

ANALYSIS OF VARIANCE OF PREREST TRIALS 11–30

Source	*df*	*MS*	*F*
Normals vs. Defectives (A)	1	49,124	40.84**
Massed vs. Distributed (B)	1	60,077	49.95**
Male vs. Female (C)	1	32,382	26.92**
A × B	1	4,416	3.67
A × C	1	3,029	2.52
B × C	1	6,843	5.69*
A × B × C	1	44	0.04
Error	152	1,202	
Between *S*s	159		
Trials (D)	19	1,988	108.60**
A × D	19	73	4.00**
B × D	19	86	1.81*
C × D	19	33	0.59
A × B × D	19	10	0.70
A × C × D	19	12	0.70
B × C × D	19	12	1.14
A × B × C × D	19	20	1.11
Error	2888	18	
Within *S*s	3040		
Total	3199		

* $P < .05$.
** $P < .01$.

Table 2 presents reminiscence scores. Table 3 summarizes an analysis of variance of reminiscence scores where reminiscence is defined as the difference between the last prerest and the third postrest score. Again, the A × B interaction is central to the hypothesis. Its absence here indi-

TABLE 2

GAIN IN PERCENTAGE OF TIME ON TARGET (REMINISCENCE) AFTER 5-MIN. REST

Subjects	*N*	Massed		Distributed	
		Mean	*SD*	Mean	*SD*
Normals					
Men	40	4.47	2.01	2.14	2.33
Women	40	5.08	1.65	2.94	3.44
Defectives					
Men	40	2.27	2.51	1.62	2.40
Women	40	3.01	3.37	0.69	2.35

TABLE 3

Analysis of Variance of Reminiscence

Source	*df*	*MS*	*F*
Normals vs. Defectives (A)	1	123.45	23.03**
Massed vs. Distributed (B)	1	138.05	25.75**
Male vs. Female (C)	1	3.75	0.69
A × B	1	5.75	1.07
A × C	1	26.80	5.00*
B × C	1	5.40	1.01
A × B × C	1	26.24	4.89*
(Error *MS*)	152	5.36	

* $P < .05$.
** $P < .01$.

cates that reminiscence in normals and defectives is similarly affected by degree of distribution of practice.

Table 4 presents a summary of an analysis of variance of the 20 postrest trials. In this instance, the A × B interaction is highly significant (.01 level) which is generated by a larger difference between the normals' curves compared to those for the defectives. Possibly, this effect could be attributed to conditioned inhibition ($_{S}I_{R}$). The greater associative capacity of the normals as well as the development of more I_R makes this interpretation tenable. The decay segment slope of the massed-practice curve for the normals is slightly steeper than that for the defectives. This trend is in accord with the prediction since it suggests a greater build-up of inhibition in the normal group. However this finding is not supported statistically as shown by the absence of an A × B × D interaction in Table 4.

TABLE 4

Analysis of Variance of Postrest Trials 1–20

Source	*df*	*MS*	*F*
Normals vs. Defectives (A)	1	30,708	1668.12**
Massed vs. Distributed (B)	1	21,923	279.79**
Males vs. Females (C)	1	11,534	147.20**
A × B	1	2,891	36.90**
A × C	1	1,300	16.60**
B × C	1	5,493	70.11**
A × B × C	1	1,981	25.29**
Error *MS*	152	78	
Between *S*s	159		
Trials (D)	19	186	2.03**
A × D	19	69	0.76
B × D	19	240	2.61**
C × D	19	20	0.22
A × B × D	19	31	0.34
A × C × D	19	27	0.30
B × C × D	19	22	0.24
A × B × C × D	19	19	0.22
Error *MS*	2888	92	
Within *S*s	3040		
Total	3199		

** $P < .01$.

Marked sex differences are evident in these analyses. In prerest practice the females within each of the four subgroups show poorer performance than the males. There are larger differences between the sexes under distributed practice (B × C interaction) and the males show greater gain with trials (C × D) independent of degree of distribution. The females show slightly larger amounts of reminiscence though this effect is not significant. (See main effect in Table 3.) The A × C interaction derives from the female normals showing more reminiscence than the normal males and a reversal of this trend within the defective group. Also, there is a reversal of this effect due to distribution of practice in the male and female defectives as compared to the trend for the male and female normals (A × B × C).

In postrest practice the females' performance, in general, falls far below that of the males with the exception of the massed defectives. In this case the performance of these *S*s is consistently above that of the defective males. The massed normal males are consistently above the females.

This reversal is mainly responsible for the A × B × C effect. The differences between massed and distributed practice were larger in the males than in the females, thus the B × C interaction. Likewise the differences between the sexes were greater in the normals than in the defectives. Probably the sex effects are attributable to a combination of motivational and motor skill factors.

Over-all these data suggest that normal *S*s build up more inhibition than do defectives although this finding is not independent of skill. This is supported by the near significant A × B interaction in prerest practice, larger amounts of reminiscence occurring in the normals, the A × B interaction in postrest performance, and the higher rate of decay trend in the massed normal group as compared to the defectives in postrest performance. This generalization requires considerable qualification, however. In pre- and postrest practice the A × B and A × B × D interactions are the only indices to differential build-up of inhibition. Such a prediction hinges upon the assumption that distribution of practice influences inhibition rather than associative mechanisms. Thus if inhibition accumulates at different rates in the normal and defective, these interaction effects could be predicted. The prerest A × B effect is near statistical significance and suggests that I_R may accumulate differentially in these *S*s, independently of performance level. The postrest effect is statistically significant but in this instance, a factor other than "temporary" inhibition seems to be operative. In the reminiscence analyses an A × B interaction is again predictable. It did not occur and the conclusion that these groups are not differentially influenced by distribution of practice is necessary. An actual decay in performance such as that seen in the massed groups in postrest practice could have provided fairly clear evidence for a differential inhibition build-up but the trend proved insignificant statistically. However, the factor of "selective sampling" lends some credibility to the hypothesis of a differential build-up of I_R. By eliminating *S*s with low performance the defective and normal groups were made more similar on the criterion measures.

On the basis of these data it seems likely that normal *S*s accumulate more inhibition than do defective *S*s. This conclusion is supported principally by the A main effect of the reminiscence analysis. However, this effect seems proportional to performance level and degree of distribution of practice. An hypothesis of a differential build-up of inhibition in these *S*s independent of performance level is not presently tenable. Further experimentation with more homogeneous groups of defective and normal *S*s should be carried out in view of trends which favor the hypothesis.

SUMMARY

The present experiment compared normals and defectives under conditions of massed and distributed practice on the rotary pursuit task. Over-all, normal *S*s showed a larger build-up of inhibition than the defectives although this effect could not be demonstrated independently of performance level and distribution of practice. There were sex differences in both prerest and postrest performance. There were some characteristics of postrest performance which could be ascribed to the development of a "permanent" inhibition effect.

REFERENCES

AMMONS, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *J. exp. Psychol.*, 1947, **54**, 263–281.

ELLIS, N. R., PRYER, M. W., & BARNETT, C. D. Motor learning and retention in normals and defectives. *Percept. mot. Skills*, 1960, **10**, 83–91.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.

(Received January 21, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 538–545

THE GENERALITY OF THE "CONDITIONING AXIOM" IN HUMAN BINARY PREDICTION [1]

PETER L. DERKS [2]

University of Pennsylvania

The linear learning models of Estes (1950) and Bush and Mosteller (1951a, 1951b, 1955) have been successful in describing the results of most binary prediction experiments. When the reward and nonreward parameters of a stochastic model are equal, the probability of the choice of an event approaches the probability of its occurrence. This "probability matching" has been reported by several investigators (Grant, Hake, & Hornseth, 1951; Hake & Hyman, 1953; Humphreys, 1939; Jarvik, 1951). However, if the nonreward parameter has no effect, the probability of the choice of the more frequent event (A) approaches one. Subjects have been found to predict A on nearly every trial (Goodnow, 1955a, 1955b; Goodnow & Postman, 1955). Thus either result is described by linear learning models with proper restrictions on their parameters.

One of the basic axioms of these models is the "conditioning axiom." In the context of the present experiments, this axiom asserts that the occurrence of an event will increase the probability that *S* will predict that event to occur again. As the event is repeated, the probability that it will be predicted increases monotonically.

To test this assertion, Estes (1957) asked *S*s to predict two nonsense syllables presented in 66:34 proportion for 2000 trials. For the last 1500 trials, the proportion of predictions of A increased monotonically as a function of three successive presentations of the syllable. However, several experiments have shown that *S*s avoid long runs in the prediction of random events (Goodfellow, 1938; Goodnow, 1955b; Skinner, 1942; Smith, 1949). Nicks (1959) reported that for event schedules of 75:25 the proportion of A predictions increased only until A had occurred twice in succession and then decreased. Anderson and Whalen (1960) obtained similar results with other schedules. These *S*s did not respond in accordance with the prediction based upon the conditioning axiom but behaved as if they believed the "gambler's fallacy" described by Jarvik (1951).

Thus, although either probability matching or 100% choice of A can be predicted from linear learning models, the changes in proportion of choice with successive occurrences of an event are not always in agreement with the monotonic increase expected from the conditioning axiom basic to the models.

EXPERIMENT I

Galanter and Gerstenhaber (1956) suggested that *S* attempts to reduce the "cost" of each prediction. If knowledge of the correct alternative is contingent on *S*'s prediction of that alternative the proportion of A choices is greater than if the correct alternative is indicated on every trial (Detambel, 1955; Goodnow, 1955a; Neimark, 1956). When *S* receives money for correct prediction the A choice proportion approaches 100% (Edwards, 1956; Galanter & Smith, 1958; Siegel & Goldstein, 1959).

---

[1] This research was supported in part by the National Institute of Mental Health Grant M-2293 and the Psychology Department of the University of Pennsylvania.

[2] Now at the College of William and Mary.

Although temporal variables are important in most learning experiments, manipulation of intertrial interval has little effect on choice proportion (Grant, Hornseth, & Hake, 1950). Outcome delay and response latency have not been examined as variables in human binary prediction. The following experiment investigated the influence of manipulations of these parameters on the proportion of choice as a function of the successive occurrence of an event and compared the results with the function asserted by the conditioning axiom.

Method

Apparatus.[3]—The equipment used in this experiment was a modification of that described by Bush and Mosteller (1955, p. 277). The *S* was seated at the panel illustrated in Fig. 1. In this study he was concerned only with the following aspects of the panel. An amber stimulus light (labeled S) on the upper panel informed *S* of the beginning of each trial. This light remained on for the duration of the trial and flashed off for 0.5 sec. indicating the end of a trial. Two white lights (labeled R and L) on the lower panel were the events. The *S* pressed one of two response buttons (labeled R′ and L′) below the event lights to predict which event would occur on a given trial. The light did not appear until after *S* had made a prediction. A latency counter (labeled LC) could register the time in 0.1 sec. between the onset of the stimulus light and the depression of the response button. This counter was operated for only one of the experimental groups.

The *E* was in a separate room where the display viewed by *S* was reproduced. All elements of the panel were the same except the response buttons which were represented by two lights. A red light on the right of *E*'s panel indicated the right response button and a blue light on the left indicated the left response button. When *S* pressed the right or left button, the corresponding light came on in *E*'s room and *E* recorded the prediction.

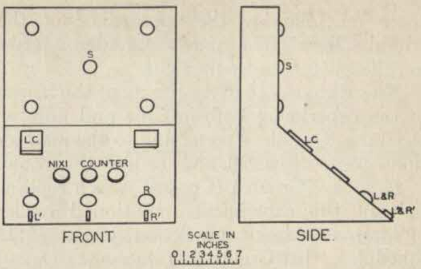


FIG. 1. Schematic diagram of human choice apparatus with S—stimulus light, L—left event light, R—right event light, L′—left response button, R′—right response button, LC—latency counter, and the nixi counter.

Time intervals were controlled by two mechanical timers. The "outcome delay" timer controlled the interval between *S*'s response and the onset of the event light. The "intertrial interval" timer controlled the duration of the event light and the subsequent onset of the stimulus light. A latency counter registered 0.1 sec. between the onset of the stimulus light and *S*'s response. Unless stated otherwise the intertrial interval and outcome delay were 1 sec. Accurate latency records were not kept as the record of *S*'s predictions required most of *E*'s attention.

The schedules of stimulus and event lights were programed on punched tapes which were stepped through two Western Union 23-A tape transmitters. A pulse from the outcome delay timer stepped the transmitter that carried the event schedule. The transmitter that carried the stimulus schedule was stepped by the intertrial interval timer or externally by *E* to position the tape. The same schedule of events was used for all *S*s. It was selected from the Rand (1955) table of a million random digits by the designation of 00 to 74 as A, the right light, and 75 to 99 as B, the left light. The series was 250 events in length and contained 187 rights and 63 lefts. There were 11 A runs, 6 trials or more in length.

Subjects.—Eighty male undergraduates at the University of Pennsylvania served as *S*s as a part of their course requirements in elementary psychology and for the monetary reward described below. They were assigned to eight groups on the basis of their order of appearance to be tested. The first 10 *S*s were Group 1, the second 10 *S*s, Group 2, etc.

Procedure.—Each *S* for all groups was instructed to guess which of two lights would come on following the appearance of the stimulus light. The guess was made by pushing the button under the light.

---

[3] The apparatus was designed and constructed under the direction of E. H. Galanter by Biophysical Instruments, Incorporated, of Germantown, Pennsylvania.

The *E* gave *S* an overestimate of the length of the experiment both in time and number of trials. Questions pertaining to the instructions were answered and *E* left the room.

Groups.—Group 1 (Control) was a replication of the experiments mentioned earlier on human binary prediction of a 75:25 schedule. In Group 2 (Outcome Delay) the outcome delay was 15 sec.; in Group 3 (Intertrial Interval) the intertrial interval was 15 sec.

In Group 4 (Restricted Latency) an attempt was made to limit the response latency. The general instructions were given with the addition that each guess was to be made "as quickly as possible" after the stimulus light flashed on. The latency counter in *S*'s room operated for this group. Since this counter, as it registered, clicked every 0.1 sec., *S* could respond to an auditory cue as well as the stimulus light. The *S* was told that he should respond in about 0.2 sec. The *S*s who required more than 0.5 sec. on over 25% of the trials were not included in the group. The average latency for the other groups was approximately 5 sec. Thus any *S* who responded at the required rate spent relatively little time on his decision. Nine *S*s failed to meet the criterion before the group of 10 was completed.

In Group 5 (Reward), *S* gained 5 cents for every correct response. The nixi counter registered once for each nickel and displayed the total amount won. The *S* was instructed concerning this tabulation with the additional reminder that "1 equals a nickel, 2 equals a dime, 5 equals a quarter, 20 equals a dollar, etc." In 250 trials these *S*s earned from $7.05 to $8.75.

In Group 6 (Risk), *S* gained 5 cents for every correct response and lost 5 cents each time he was wrong. The nixi counter displayed the net total. The *S* was instructed concerning this tabulation plus the same reminder given to Group 5. These *S*s earned from $3.25 to $5.50 in 250 trials.

In Group 7 (Contingent) the same 75:25 schedule was used, but the event light only appeared if it was predicted correctly. The *S* was instructed:

> Try to push the correct button as often as possible. These buttons are independent of each other. There are trials on which both buttons are correct, trials on which neither button is correct, and trials on which only one button is correct. Nevertheless, you may push only one button per trial. When a correct button is pushed the light above it will come on. Even if both buttons are correct only the light over the button you push will come on. Thus you only receive information about one button at a time.

In Group 8 (Contingent-Risk) the 75:25 schedule was again used with the contingent procedure described for Group 7, but *S* received 5 cents when correct and lost 5 cents when wrong with the total presented on the nixi counter. These *S*s earned from $3.20 to $5.70.

Results

Probability matching.—Differences in probability matching are usually determined from the average of the final choice proportions of a group of *S*s. However it would be possible for 1 *S* to predict A half the time on a 75:25 schedule while another predicted that event on every trial. The average data for the 2 *S*s would be a probability match while the individual results would not. To overcome this source of error each group was also described in terms of individual *S*s who were probability matchers (PMs).

A PM was defined as any *S* whose proportion of A predictions on the last 100 trials was within ±2 *SD* of the actual proportion of that event. The *SD* estimated by:

$$\hat{\sigma} = \sqrt{N(\pi_1 \cdot \pi_2)} \qquad [1]$$

where $\hat{\sigma}$ = estimated *SD*, π_1 and π_2 = the proportion of A and B, and N = number of trials.

For a 75:25 schedule 2 *SD* = approximately 9. Thus any *S* who predicted A from 66 to 84 times on the last 100 trials was considered to be a PM. The number of PMs and the mean number of A predictions on the last 100 trials for each group is summarized in Table 1.

Extension of the intertrial interval or the outcome delay to 15 sec. is not sufficient to significantly alter choice proportion or number of PMs. Group 1 (Control), Group 2 (Outcome Delay), and Group 3 (Intertrial Interval) are not significantly different. These results agree with other binary prediction experiments in which intertrial interval was a variable.

However Group 4 (Restricted Latency), when compared with Group 1, has significantly more PMs ($P < .05$, $B = 9$, $D = 4$) from the Finney table of the Fisher test of χ^2 (Siegal, 1956) and a lower proportion of A predictions on the last 100 trials ($P < .05$, $t = 2.45$, $df = 18$). Furthermore, of the 9 *S*s who were not included in this group because they failed to respond within 0.5 sec. on 25% of the trials, only 3 were PMs. Therefore, speed instructions resulted in an increase in PMs if *S* responded to the instructions by making decisions at the average rate of approximately .5 sec. per decision (the criterion for inclusion in Group 4), but did not result in an increase of PMs in the case of *S*s who responded to the speed instructions by speeding up their decisions (as compared with Group 1, which worked at the rate of approximately 5 sec. per decision), but only to an average rate between .5 and 3.0 sec. per decision.

When incentive is manipulated, Group 6 (Risk) has the highest proportion of A predictions, followed in order by Group 5 (Reward) and Group 1. These results are in the direction reported in other studies but the differences here are not significant. Although *S*s in Groups 5 and 6 all reported that they believed they were going to be paid they did not significantly increase their choice proportion over *S*s who received no money. Numbers that registered on a counter did not affect behavior as did actual nickels or chips for money of other studies (Bush & Mosteller, 1955, Ch. 13; Edwards, 1956; Galanter & Smith, 1958; Siegel & Goldstein, 1959). The reason for this difference is not clear.

TABLE 1

PROBABILITY MATCHERS (PMs) AND PROPORTION OF MORE FREQUENT EVENT PREDICTIONS ($\overline{P}_1$) ON FINAL 100 TRIALS

Group	Condition	No. of PMs	$\overline{P}_1$
1	Control	4	.850
2	Outcome Delay	5	.820
3	Intertrial Interval	5	.885
4	Restricted Latency	9	.764
5	Reward	3	.879
6	Risk	3	.880
7	Contingent	5	.803
8	Contingent-Risk	1	.948

The *S*s in the contingent groups (Groups 7 and 8), when questioned after the experiment, all said they accepted the independence of the events. Verbal estimates of "proportion correct" for each event did not sum to 1 for 13 of the 20 *S*s. Since none of these groups had more than 5 PMs out of 10 *S*s, the differences between number of probability matchers are not significant. However, Group 8 (Contingent-Risk) had a significantly greater proportion of A predictions than Group 1 ($P < .05$, $t = 2.48$, $df = 18$) and Group 7 (Contingent) ($P < .01$, $t = 3.86$, $df = 18$).

These results agree with studies which reported that a contingent-risk procedure resulted in a greater A choice proportion than other procedures. However, in this study, the contingent procedure did not result in a significant difference in choice proportion from a noncontingent one when *S*s were unrewarded.

Choice and event repetition.—The proportion of choice as a function of successive occurrence of an event was calculated from *S*s' responses to sequences of six or more repetitions of

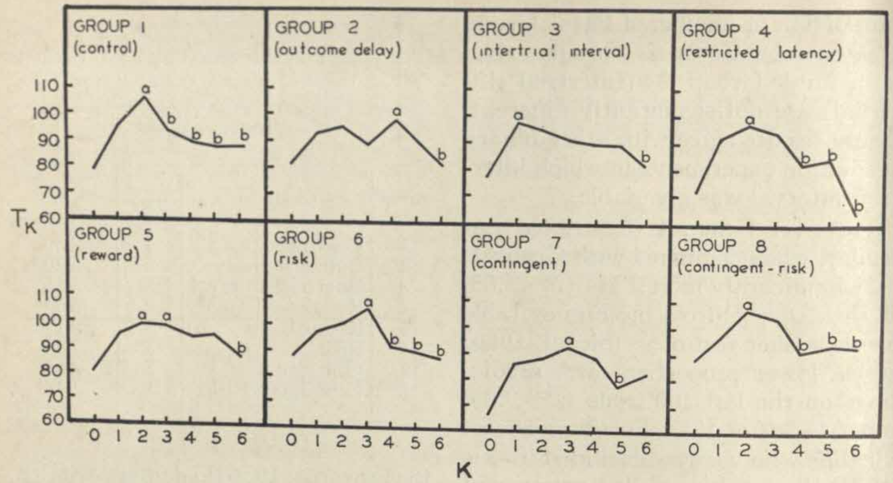


FIG. 2. T_k as a function of k in Exp. I: a—Maximum T_k; b—Subsequent $T_k > 2$ *SD* below maximum.

A.[4] There were 11 of these runs in the 250-trial schedule. The predictions of A after it had occurred 0, 1, 2, . . ., 6 times in a row were summated over the 10 *S*s in each group for the 11 sets of six successive A events. Thus the sample consisted of 110 cases in which A followed B either 0, 1, 2, . . ., 6 times in succession in the sets of A runs that were six or longer. This is:

$$T_k = \sum_{i=1}^{10} \sum_{j=1}^{11} n_{kji} \qquad [2]$$

where:

i = any individual subject ($i = 1, 2, \ldots, 10$)

j = any set of six successive occurrences of A ($j = 1, 2, \ldots, 11$)

k = any number of successive occurrences of A within a succession of six or more such occurrences ($k = 0, 1, \ldots, 6$)

n_{kji} = number of A predictions following k successive occurrences of A for set j and subject i ($n = 0, 1$)

T_k = total number of A predictions following k successive occurrences of A for all 11 sets and all 10 *S*s (T_k is a positive integer, $0 \leq T_k \leq 110$)

Figure 2 summarizes T_k as a function of k for all groups. The conditioning axiom states that T_k will increase monotonically as k increases, i.e., the predictions of A will increase with its successive occurrence. If there is a T_k more than 2 *SD* less than any preceding T_k the function cannot be described as monotonically increasing. Equation 1 gives an estimate of 2 *SD* for T_k of approximately 9. In Fig. 2 the maximum T_k and any subsequent T_k that is 2*SD* less than the maximum is noted. T_k is not a monotonically increasing function of k for any group.

EXPERIMENT II

Edwards (1961) found that for schedules of 1000 trials duration "the asymptotic probability of prediction was more extreme than the probability of occurrence of the event predicted." He computed the proportion of choice of an event as a function of its successive occurrence using the method described by Nicks (1959). On the first 400 trials the function rises to a peak value after which the proportion of predictions of A decreases. However, on the last 600 trials the proportion of choice of an event increases monotonically as a function of the successive occurrence of that event as asserted by the conditioning axiom. Lindman and Edwards (1961) have found that this change is already underway after 100 trials with a 50:50 schedule. As mentioned earlier Estes (1957) found that the function increased monotonically on Trials 500 to 2000.

---

[4] This statistic was suggested by R. R. Bush and H. C. Morlock of the University of Pennsylvania in an unpublished memorandum.

Experiment II investigated the proportion of choice of A as a function of its successive occurrence in a 1000-trial sequence. In order to examine any progressive change in this function T_k was computed for successive blocks of 250 trials.

Method

The same equipment described in Exp. I was used. Ten male undergraduates (Group 9) at the University of Pennsylvania were paid $4.00 each to serve as *S*s. They were given the same instructions cited in Exp. I. The length of the experiment was given as "1000 trials and about 2 hr." Each *S* predicted the same 75:25 noncontingent schedule described in Exp. I. The 250 trials of this schedule were repeated four times without interruption to make 1000 trials.

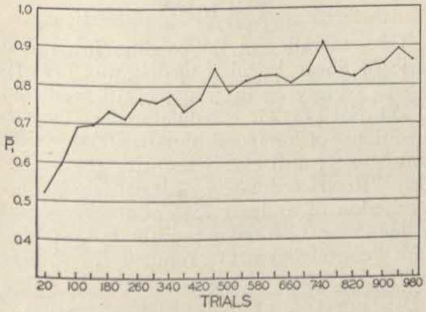


FIG. 3. Proportion of predictions of the more frequent event on a 75:25 schedule ($\bar{P}_1$) as a function of trials in blocks of 40.

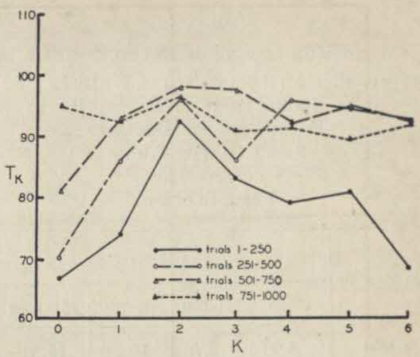


FIG. 4. T_k as a function of *k* in Exp. II.

Results

The proportion of predictions of A as a function of trials in blocks of 40 is presented in Fig. 3. On this 75:25 contingent schedule the choice proportion continued to increase throughout 1000 trials. However, 3 *S*s were still PMs on the last 100 trials. Although group choice proportion does not stabilize in 1000 trials, individual *S*s may consistently probability match.

T_k as a function of *k* is summarized for each quarter of the sequence in Fig. 4. On Trials 1–250, T_k is the "peaked" function of *k* found in the earlier experiments of this study. A curve that increases monotonically can fit T_k as a function of *k* for all but one point for Trials 251–500 (T_k for $k = 3$ is more than 2 *SD* below the peak value at $k = 2$) and all points for Trials 501–750. However, T_k is a linear function of *k* for Trials 751–1000 and the slope is approximately zero.

In order to examine the cause of the change in T_k, some of the information reflected by T_k was tabulated more fully. The mean run length of predictions of A and the number of predictions of B following its occurrence are summarized in Table 2. The run length of A pre-

TABLE 2

MEAN RUN LENGTH OF PREDICTIONS OF EVENT A ($\overline{R}_1$) AND TOTAL NUMBER OF PREDICTIONS OF EVENT B FOLLOWING ITS OCCURRENCE (r_2) FOR FOUR SUCCESSIVE 250-TRIAL BLOCKS

Prediction	Trials			
	1–250	251–500	501–750	751–1000
$\overline{R}_1$[a]	3.93	5.70	10.20	17.51
r_2[b]	228	184	155	91

Note.—Event B occurred 630 times for the 10 *Ss* in each 250-trial block. All *t* values were obtained by the Lord technique as described by Mosteller and Bush (1954).

[a] For $\overline{R}_1$, Trials 751–1000 differ from 1–250 ($P < .02$, $t = 3.15$, $df = 9$) and 251–500 ($P < .05$, $t = 2.64$, $df = 9$), and Trials 1–250 differ from 501–750 ($P < .02$, $t = 2.75$, $df = 9$). None of the other differences are significant.

[b] For r_2, Trials 751–1000 differ from 1–250 ($P < .01$, $t = 3.88$, $df = 9$) and 251–500 ($P < .02$, $t = 2.80$, $df = 9$). None of the other differences are significant.

dictions increases from quarter to quarter while the B predictions that follow A's occurrence decrease. The tendency to predict longer runs is reflected in T_k values that remain nearly constant for $k \geq 2$. Since *Ss* also predict B to repeat less often for successive blocks of 250 trials, T_k for $k = 0$ also increases with each block of 250 trials. By Trials 751–1000, T_k is constant over k from 0 to 6. The monotonic increase in T_k as a function of k on Trials 251–500 and 501–750 is caused by the gradual increase of T_k for $k = 0$ compared with the rapid increase for $k = 6$. The *S* predicts longer A runs sooner than he ceases to predict repetitions of B. As a result, T_k increases monotonically as a function of k and briefly fulfills the assertion of the conditioning axiom.

DISCUSSION

The results of these experiments show that the assertion of the conditioning axiom that is basic to linear learning models does not fit the fine grain of the data in many types of human binary choice experiments. Variations in temporal parameters, incentive parameters, and the contingency of the events on *S*'s response altered probability matching but did not result in a proportion of A predictions that increased monotonically as a function of the successive occurrence of the event. In every case *Ss* behaved as if they believed the gambler's fallacy, i.e., they did not predict very long or very short runs.

When the experiment was extended to 1000 trials *Ss* began to predict longer runs for successive 250-trial blocks. As a result the proportion of A predictions was the monotonically increasing function of its successive occurrence predicted by the conditioning axiom.

On the final 250 trials *Ss* predicted fewer repetitions of B than in the other blocks of 250 trials. These results agree with those of Goodnow and Pettigrew (1955) and Anderson (1960) who found that *S* will repeat predictions of events that tend to repeat and will not predict repetitions of events that do not tend to repeat. A is predicted to continue to occur or to follow the occurrence of B; B is predicted only after long runs of A.

The monotonically increasing function of T_k against k which is found on Trials 251–750 is a result of the comparatively rapid development of prediction of long runs of A with respect to the decrease in the prediction of repetitions of B. It must be concluded that, in general, the assertion of the conditioning axiom does not give an accurate description of human binary prediction.

SUMMARY

The generality of the conditioning axiom basic to linear learning models was investigated. Ninety *Ss* in groups of 10 predicted a 250-trial 75:25 schedule under various conditions of temporal duration, incentive, contingency, and experience with the schedule. "Restricted latency" instructions plus a criterion of at least 75% of the responses within .5 sec. resulted in probability matching and "contingent-risk" resulted in nearly 100% prediction of the more frequent event (A). However, a sample of the proportion of prediction of A (T_k) as a function of its successive occurrence (k) did not show the monotonically increasing function asserted by the conditioning axiom.

When *S*'s experience with the schedule was extended to 1000 trials, T_k did increase monotonically as a function of *k* for Trials 251 to 750. The monotonic increase was the result of the differential rate at which *S*s began to predict long runs of A and decreased the prediction of repetitions of B. It was concluded that the assertion of the conditioning axiom does not describe most human binary prediction.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

ANDERSON, N. H., & WHALEN, R. E. Likelihood judgments and sequential effects in a two-choice probability learning situation. *J. exp. Psychol.*, 1960, **60**, 111–120.

BUSH, R. R., & MOSTELLER, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323. (a)

BUSH, R. R., & MOSTELLER, F. A mathematical model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423. (b)

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

DETAMBEL, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, **49**, 97–105.

EDWARDS, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

EDWARDS, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

GALANTER, E. H., & GERSTENHABER, M. On thought: The extrinsic theory. *Psychol. Rev.*, 1956, **63**, 218–227.

GALANTER, E. H., & SMITH, W. A. S. Some experiments on a simple thought problem. *Amer. J. Psychol.*, 1958, **71**, 359–366.

GOODFELLOW, L. D. A psychological interpretation of the results of the Zenith Radio experiments in telepathy. *J. exp. Psychol.*, 1938, **23**, 601–632.

GOODNOW, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116. (a)

GOODNOW, J. J. Response sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 624–630. (b)

GOODNOW, J. J., & PETTIGREW, T. F. Effect of prior patterns of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, **49**, 381–389.

GOODNOW, J. J., & POSTMAN, L. Probability learning in a problem-solving situation. *J. exp. Psychol.*, 1955, **49**, 16–22.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

GRANT, D. A., HORNSETH, J. P., & HAKE, H. W. The influence of the intertrial interval on the Humphreys' "random reinforcement" effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, **40**, 609–612.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

HUMPHREYS, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

LINDMAN, H., & EDWARDS, W. Supplementary report: Unlearning the gambler's fallacy. *J. exp. Psychol.*, 1961, **62**, 630.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954.

NEIMARK, E. D. Effects of types of non-reinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

NICKS, D. C. Predictions of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

RAND CORPORATION. *A million random digits with 100,000 normal deviates.* Glencoe, Ill.: Free Press, 1955.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

SKINNER, B. F. The processes involved in the repeated guessing of alternatives. *J. exp. Psychol.*, 1942, **30**, 495–503.

SMITH, M. H. Spread of effect is the spurious result of non-random response tendencies. *J. exp. Psychol.*, 1949, **39**, 355–368.

(Received March 4, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 546–554

MULTIDIMENSIONAL SCALING OF FACIAL EXPRESSIONS

ROBERT P. ABELSON AND VELLO SERMAT

Yale University

The ability to identify the emotions displayed in facial expressions has long been a subject of psychological study. The early literature, reviewed by Woodworth (1938), focused attention on the surprisingly frequent inability of untrained *S*s to report the emotion intended by the stimulus. Woodworth pointed out that errors of identification usually fell into categories psychologically similar to the intended one. A long struggle to organize these similarities in some coherent way culminated in Schlosberg's (1952, 1954) proposals of a two-dimensional and later a three-dimensional theory of facial expressions. In studies by Engen and Levy (1956), Engen, Levy, and Schlosberg (1957, 1958), and Triandis and Lambert (1958) it was shown that judges can reliably locate posed facial expressions along each of Schlosberg's three scales: pleasant-unpleasant, attention-rejection, and sleep-tension. Any given emotional expression can thus be represented by three coordinates, the respective locations on the three component scales. Engen, Levy, and Schlosberg (1958) have given these coordinates for 48 photographs (the "Lightfoot Series") portraying a variety of situations, moods, and emotions via the face of a young actress; e.g., maternal love, anxiety that her plane will crash, savoring a Coca-Cola, etc. The coordinates of each stimulus are the median ratings by 96 male undergraduates on three nine-point scales.

Generally speaking, facial expressions with similar coordinates on Schlosberg's three dimensions are susceptible to confusion with one another, e.g., joy with plasant surprise, hatred with disgust, etc. This elucidates the classical finding of frequent misidentification, and also exemplifies the general relationship between psychological similarity and the probability of "intrusion errors."

However, the particular three dimensions used by Schlosberg are not necessarily the only dimensions or the best dimensions for explaining confusion data or for general usage in this research area. There is the possibility that one or more of Schlosberg's scales, while understandable when made explicit to judges, are unimportant in uninstructed perception of facial expression; or conversely, that one or more important scales have been omitted. The Schlosberg theory thus suffers from the weakness that no direct account is taken of the dimensions operative for the judges; *E* imposes particular dimensions of his own choosing and is arbitrarily forced to give them equal weight.

This vulnerable aspect of the Schlosberg system arises because the historical development of the system was quite indirect. There is only a very loose logical correspondence between confusion data on the one hand and scale locations on the other. A direct correspondence can best be obtained by applying a single analytical operation to a set of confusion data to derive a multidimensional set of scale values. Shepard's (1957) scaling procedure represents one way to achieve this. An alternative method is detailed below.

A few attempts have already been made to derive scale values directly from confusion or similarity data. Unfortunately, these studies for one reason or another are not directly applicable to the present problem. Hulin and Katz (1935) made an early attempt, but since sophisticated analytical methods were not available at the time, their data were not gathered in the most useful way. Nummenmaa and Kauranne (1958) applied Ekman's (1954) factoring technique to judgments of similarity of facial expressions; however, Ekman's technique spuriously increases the dimensionality of the solution and does not produce clear results, as surmised by Messick and Abelson (1957) and demonstrated empirically by Mellinger (1958). Osgood (1955) produced an interesting three-dimensional facial system mediated by verbal labels attached to poses. The use of verbal mediation and other complicating factors make this study difficult to interpret in the present context.

The multidimensional scaling techniques developed recently by Torgerson (1952) and others permit investigation of a stimulus domain *without preselection of component dimensions.* Judges are asked to respond to stimuli in terms of their similarity to each other. From these judgments, estimates of the "distances" between stimuli are derived. Coordinates of the stimuli in a psychological space of appropriate dimensionality are obtained by a matrix factoring procedure. The identification of the dimensions is post hoc, as in factor analysis.

The present paper reports an application of multidimensional scaling to the domain of facial expressions. Here, the existence of Schlosberg's scales is very useful, for it allows better interpretation of the dimensions derived from the multidimensional scaling operation.

METHOD

Subjects.—The *S*s were 30 women graduate and special students residing at a women's dormitory in New Haven, Connecticut. Ages ranged from 20–55 yr., with a median of 23 yr. The *S*s were solicited by posting an advance notice briefly indicating the topic of study and requesting volunteers at designated times. Three groups of 10 *S*s each were used.

Procedure.—Thirteen photographs had been selected from the Lightfoot Series[1] to tap collectively the full range of each of Schlosberg's three dimensions. All 78 possible pairs of these stimuli were presented to *S*s in a single random order. The stimulus pairs were projected onto a screen by means of a pair of slide projectors. Each individual stimulus appeared 12 times, 6 times on the left and 6 on the right. Each stimulus pair was exposed for 6 sec., during which time each *S* judged the dissimilarity of the members of the pair and recorded a number from 1 to 9 on a response sheet. The *E* then removed the slides, readied and announced a new pair, and shifted the new slides into position simultaneously. This pause between presentations lasted 10 sec. Pairs were designated on *S*'s response sheets by code letters.

The judgment procedure was preceded by a warm-up period during which the rating scale was explained and three pairs of stimuli were shown as illustrations. The warm-up stimuli were Lightfoot Faces different from the 13 crucial stimuli.

Instructions.—The rating scale instructions to *S*s were as follows:

> I am going to show you a number of photographs of the same person, and each time I will present two of these pictures side by side. Your task will be to judge how much the two pictures differ in their emotional expression. I would prefer if you try to make a rather spontaneous judgment, without trying to theorize too much as to what emotion is being expressed and how this should relate to the emotional expression of the other picture. Rather you should think in about these terms: emotional expressions of people differ, some being more similar than others. If we should give the difference a numerical score, we would quite arbitrarily decide that we will give the score of 1 to a pair of two facial expressions which are most alike in their emotional content; a score of 9 to two expressions which are furthest apart, and a score of 5 to a pair of pictures which fall about half-way between these two extremes.
>
> Just to give you some idea what I have in mind, I will now show you a few examples. This first pair of pictures portrays

---

[1] The authors are grateful to Harold Schlosberg for generously providing the necessary slides.

emotional expressions which can be said to be fairly similar; you might score this pair 1 or 2 or 3. . . . The next pair shows facial expressions which are close to the extreme opposites, so you might score them from 7 to 9. . . . This third pair fits in somewhere between the two preceding pairs, so you might score it 4 or 5 or 6. . . . Are there any questions?

Now would you please look at your sheets. There are spaces left for 78 entries, each entry consisting of one score between the values 1 and 9 for one pair of faces. The same pictures will occur again and again, but in different combinations, and you are not rating single pictures but rather the difference between any two pictures at a given time. Each time when I show a pair, I call out two letters, to indicate which entry on the sheet this pair refers to. This is to avoid errors in writing scores into their proper locations on the sheet. Please check each time that the two letters I named are exactly the same as the letters before the empty space where you are going to write in your score for this pair.

Please try to write in your judgments as quickly as possible; thinking too long will only confuse you. It is also absolutely essential that you do not miss *any* single entry; even if it is difficult to judge some pictures, put down your first best guess. Leaving some entries empty will create much more serious problems for my reresearch than being uncertain about the correct score and putting in the best guess. Also, try to use the whole scale of numbers, that is, use every number from 1 to 9 at some time or other. And do not go back to check what you have written down before. Just go on from one pair to another. And of course, please avoid influencing one another, or discussing, or exchanging any opinions or advices throughout this time.

Are there any further questions? Fine, then we will start with the first pair of facial expressions. Please look at it and then put down your best estimate by writing a number into the first entry space.

Remember: rating 1 indicates that the two faces differ very, very little in their emotional expression, rating 9 indicates that they differ extremely much, and the other numbers represent gradations of difference in between.

The entire judgment session lasted approximately 30 min. The data consisted of the dissimilarity ratings by 30 *S*s of 78 stimulus pairs. The data were later punched into IBM cards for the subsequent analysis.

RESULTS

Table 1 reproduces from Engen, Levy, and Schlosberg (1958) the scale values of the 13 stimuli on the three Schlosberg scales. One notes in this table a strong positive correlation between the values on Dimensions

TABLE 1

SCALE VALUES ON THE THREE SCHLOSBERG DIMENSIONS FOR 13 FACES FROM THE LIGHTFOOT SERIES

Stimulus	Pose	(V_1) Pleasant-Unpleasant	(V_2) Attention-Rejection	(V_3) Tension-Sleep
7	Grief at death of mother	3.8	4.2	4.1
13	Savoring a Coke	5.9	5.4	4.8
15	Very pleasant surprise	8.8	7.8	7.1
16	Maternal love—baby in arms	7.0	5.9	4.0
20	Physical exhaustion	3.3	2.5	3.1
28	Anxiety—something is wrong with her plane	3.5	6.1	6.8[a]
29	Anger at seeing dog beaten	2.1	8.0	8.2
30	Physical strain—pulling hard on seat of chair	6.7	4.2	6.6
32	Unexpectedly meets old boy friend	7.4	6.8	5.9
36	Revulsion	2.9	3.0	5.1
37	Extreme pain	2.2	2.2	6.4
51	Knows her plane will crash	1.1	8.6	8.9
56	Light sleep	4.1	1.3	1.0

[a] The entry 8.2 in the original published table was a misprint.

TABLE 2

Squared Distances, d^2, between Stimuli, as Derived from Dissimilarity Ratings

Stimulus	7	13	15	16	20	28	29	30	32	36	37	51	56
7	0.00	4.05	8.25	5.57	1.15	2.97	4.34	4.90	6.25	1.55	1.68	6.57	3.93
13	4.05	0.00	2.54	2.69	2.67	3.88	8.53	1.31	1.88	4.84	5.81	7.43	4.51
15	8.25	2.54	0.00	2.11	8.98	9.27	11.87	2.56	0.74	9.25	7.92	8.30	8.47
16	5.57	2.69	2.11	0.00	3.78	6.05	9.78	4.21	0.45	4.92	5.42	8.93	3.48
20	1.15	2.67	8.98	3.78	0.00	2.34	7.12	5.90	4.77	2.22	4.34	8.16	1.60
28	2.97	3.88	9.27	6.05	2.34	0.00	1.36	5.18	5.45	4.17	4.72	4.66	4.89
29	4.34	8.53	11.87	9.78	7.12	1.36	0.00	8.47	10.20	5.44	4.31	1.57	9.18
30	4.90	1.31	2.56	4.21	5.90	5.18	8.47	0.00	2.63	5.45	3.79	6.49	6.05
32	6.25	1.88	0.74	0.45	4.77	5.45	10.20	2.63	0.00	7.10	6.58	9.77	6.55
36	1.55	4.84	9.25	4.92	2.22	4.17	5.44	5.45	7.10	0.00	1.98	4.93	4.12
37	1.68	5.81	7.92	5.42	4.34	4.72	4.31	3.79	6.58	1.98	0.00	4.83	3.51
51	6.57	7.43	8.30	8.93	8.16	4.66	1.57	6.49	9.77	4.93	4.83	0.00	12.65
56	3.93	4.51	8.47	3.48	1.60	4.89	9.18	6.05	6.55	4.12	3.51	12.65	0.00

V_2 and V_3. This is not an accident of selection for the present study of 13 faces out of the total set of 48, since the correlation between these two dimensions in the total set is even higher.

The configuration of stimuli in Schlosberg's three-coordinate space is therefore quite flat; i.e., a plane is not a bad fit to the stimulus points. This fact becomes important later.

For each stimulus pair, a distribution of dissimilarity ratings was available from the present study. The mean of this distribution is a possible estimate of the "distance" between the members of the pair; however, a more sensitive estimate can be obtained from the method of successive intervals. Diederich, Messick, and Tucker's (1957) version of this method was used,[2] yielding estimates of the distances between each stimulus and every other. These distances are determined on an interval scale only; a zero point on the distance scale was located by Messick and Abelson's (1956) procedure. Table 2 presents the *squares* of the resulting distances between each stimulus and every other. The self-distance of each stimulus is zero.

Before proceeding to the multidimensional analysis of the distance matrix, it is possible to make comparisons between the present distance data and the previous scale data. Suppose one assumes that the interstimulus distances can be explained by interstimulus differences on the Schlosberg scales; i.e., that a three-dimensional space composed of the Schlosberg dimensions locates the stimuli so as to generate the distance data.

In setting up such a comparative analysis, the following considerations were employed:

1. Euclidean geometry requires that the squared distance between Stimuli j and k equal the sum of the squared scale differences between Stimuli j and k on each of the three component dimensions.
2. However, since the ranges of the Schlosberg scales are not calibrated one with another, the three component dimensions might have to be weighted differentially.
3. One must realistically expect the interstimulus distances to depend at least to a minor extent upon a variety of "extraneous"

---

[2] Carl Helm of the Educational Testing Service graciously performed this analysis.

dimensions. Each such dimension will *increase* the interstimulus distances beyond what one would predict from the Schlosberg scales. This variable error can be partitioned conceptually into a constant error and a random error. The latter category also serves to include sampling and measurement errors in the distance data.

4. The Schlosberg scale values are assumed to have been measured without error.

The form of the assumed relationship is thus:

$$d_{jk}^2 = b_1(V_{1j} - V_{1k})^2 + b_2(V_{2j} - V_{2k})^2 + b_3(V_{3j} - V_{3k})^2 + C + e_{jk} \quad [1]$$

for all $j \neq k$, where the d_{jk}^2 are the distance data of Table 2, the quantities $(V_{1j} - V_{1k})^2$, etc. are computed from the scale data of Table 1, the b's and C are unknown constants, and the e_{jk} are errors with zero mean and a "random" distribution.

The form of the problem is that of a multiple regression with an N of 78 stimulus pairs, each pair providing a criterion observation d_{jk}^2 and three predictor observations of the form $(V_j - V_k)^2$. The unknown constants were determined by the usual least squares procedure, and the results are presented in the first column of Table 3.

The very high multiple R associated with the regression indicates that the Schlosberg scales predict three-quarters of the variance in the distance data. The regression weights suggest a particular calibration of the three scales (an unachieved goal often mentioned in the literature). It follows from Equation 1 that if the scales are to have equal weight in predicting interstimulus distance the *values* on each scale must be multiplied by a factor proportional to the square root of the appropriate b. A rough indication of these factors will suffice, since the present modest collection of data is not the last word in the matter. Using all three scales, a rough calibration is provided by the factors 4:3:2. In other words, if pleasant-unpleasant is located on a scale from 1 to 9, then tension-sleep should be located on a scale from 1 to 7, and attention-rejection on a scale from 1 to 5.

TABLE 3

WEIGHTS TO APPLY TO THE SCHLOSBERG SCALES TO BEST APPROXIMATE THE DISTANCE DATA ACCORDING TO EQUATION 1

Scale	V_1, V_2, & V_3	V_1 & V_3	V_1 & V_2	Best Single Scale
V_1	.156	.157	.148	.151
V_2	.034	—	.088	—
V_3	.091	.112	—	—
Constant error, C	1.72	2.16	2.53	3.49
Multiple R	.868	.854	.811	.706

Table 3 also gives the results for two-scale regressions, omitting one scale from the predictive formula. Omission of attention-rejection does not lower prediction appreciably; omission of tension-sleep is slightly more damaging. The reason attention-rejection is less helpful in prediction is not because it is a poor scale, but rather because of redundancy: squared scale differences on this scale correlate *positively* with squared scale differences on pleasant-unpleasant, whereas squared scale differences on tension-sleep correlate *negatively* with squared scale differences on pleasant-unpleasant.

Considering only two scales at a time, the rough calibration factors are: for V_1 and V_3, 5:4; for V_1 and V_2, 4:3.

The last column of Table 3 indicates that the best single contribution to prediction is, not surprisingly, made by the pleasant-unpleasant dimension, which accounts for half the variance in the distance data.

The next step was to derive a multidimensional set of stimulus coordinates from the distance data with-

TABLE 4

Multidimensional Scaling Results: Coordinates of the Stimuli on Five Dimensions

Stimulus	Dimension				
	I	II	III	IV	V
7	−.65	−.64	.24	−.31	−.24
13	.85	.06	−.36	−.71	−.26
15	1.64	1.02	.22	.19	.06
16	1.15	−.16	−.05	.87	−.14
20	−.19	−1.14	−.48	−.11	−.51
28	−.84	−.01	−1.09	−.17	.28
29	−1.74	.77	−.48	.30	.43
30	.73	.54	.42	−.90	.37
32	1.43	.32	−.40	.26	−.04
36	−.72	−.55	.65	.10	−.71
37	−.66	−.23	1.05	.10	.51
51	−1.23	1.55	.21	.16	−.39
56	.22	−1.52	.06	.23	.64
Sum of squares	14.02	8.83	3.80	2.50	2.14

out reference to the Schlosberg scales. The details of the standard procedure are given by Torgerson (1958).

The dimensionality of the solution is determined from the pattern of latent roots of a Matrix B derived from the matrix of squared distances. For the present data, five dimensions were required. The coordinates of the stimuli on these five orthogonal dimensions are given in Table 4. The row labeled "sum of squares" gives the sum of squared distances of stimulus points from the origin on each dimension. This quantity may be thought of as the variance or "spread" accounted for by each dimension. It is seen that the first two dimensions are far more important than the latter three, accounting between them for 73% of the total spread. The factoring procedure could have been stopped after two dimensions to provide a very crude solution.

The dimensions are initially unidentified. As in factor analysis, a rotation of the coordinate axes may be necessary to yield psychologically interpretable dimensions, although the "simple structure" criterion of factor analytic rotation is not necessarily relevant. In the present case, the Schlosberg scales provide useful guidance in interpretation of dimensions. Table 5 gives the correlations between coordinates on the five present dimensions and values on the three Schlosberg scales. It should be remembered that two different samples of judges were involved, differing in age and sex. The judgment task itself was different in that no reference was ever made in the present task to any explicit dimensions.

In view of this, the correlations of .947 between Dimension I and pleasant-unpleasant, .878 between Dimension II and attention-rejection, and .917 between Dimension II and tension-sleep are extremely high. The multiple correlations, .962, .982, and .965, given in the last column of Table 5 show that the each of the Schlosberg scales can be located without appreciable error by particular orientations of the Axes I–V. None of the Schlosberg scales, therefore, can be called trivial in the sense that the stimulus separations it provides are not to be found in direct judgments of dissimilarity. However, one again encounters redundancy between Scales V_2 and V_3; Dimension II is representative of both attention-rejection and tension-sleep. Thus while there are

TABLE 5

Correlations between the Present Five Dimensions and the Three Schlosberg Scales

Schlosberg Scale	Dimension					Multiple R (all five)
	I	II	III	IV	V	
V_1	.947	.121	−.094	−.031	.061	.962
V_2	.026	.878	−.374	.210	.094	.982
V_3	−.295	.917	.011	−.041	.046	.965

three scales and five dimensions, there are essentially three extra, uninterpreted dimensions rather than two.

Attempts were made to give psychological identification to the three extra dimensions. One may seek rotations of Axes III, IV, and V of Table 4 yielding meaningful interpretations. Alternatively, a somewhat more sophisticated approach would be as follows: (*a*) locate a new axis I′ correlating maximally with V_1; (*b*) locate a new axis II′ correlating maximally with an optimal combination of V_2 and V_3; (*c*) solve for a triplet of axes, III′, IV′, and V′ orthogonal to I′ and II′; and (*d*) rotate III′, IV′, and V′ to meaningful positions. Both approaches were tried, but neither met with sufficient success to justify pronouncements on the nature of the extra dimensions. The subjective verdict was that these dimensions were "hash."

DISCUSSION

The present study provides evidence that the Schlosberg scales are sound: they predict the dissimilarity data fairly well, and the extra dimensions that appear in the multidimensional scaling operation do not suggest new meaningful scales to replace or supplement the old ones.

The Schlosberg scales do suffer, however, from redundancy in that two scales seem to serve as well as three. If a choice must be made between tension-sleep and attention-rejection, the former deserves a very slight preference. As a supplement to the pleasant-unpleasant scale, the tension-sleep scale is better than the attention-rejection scale in predicting the dissimilarity data. Other relative demerits of the attention-rejection scale have been noted elsewhere: difficulty in explaining the scale to *Ss* (Schlosberg, 1952); lower inter- and intrarater reliability (Engen et al., 1958); lower stability across scaling methods (Engen & Levy, 1956). There is of course the possibility of devising a new scale which compromises between the two scales in question.

Osgood (1960, p. 362) reports a study done with Hastorf in which facial expressions were rated on scales of *pleasantness* and *activation*. This two-dimensional scheme, apparently a selective refinement from Osgood's (1955) earlier three-factor system, is probably close to optimal. The earlier three-dimensional system is hard to interpret, since it is based on different poses by different actors, thus confounding inter- and intraindividual dimensions of expression.

There are several possible explanations for the uninterpretibility of the three extra dimensions of the present study. Conceivably, sensible interpretations were overlooked. Alternatively, it may be that the emergence of these dimensions was somehow an artifact of the method of dealing with the distance data.

There are three possibilities here: that the assumption of a Euclidean geometry is to some extent inappropriate; that there is a "communality problem," i.e., that each stimulus introduces a small dimension of its own; that some monotone transformation of the distance scale would produce cleaner results. There is little definitive knowledge of how to attack the first two problems, but on the third problem, it has been suggested by Helm, Messick, and Tucker (1959) that an exponential transformation of the distance scale can often reduce dimensionality. Such a transformation was attempted with the present data, but the outcome was unsatisfactory. Instead of reducing the dimensionality, the revised solution additionally produced a sizeable negative root in the factoring process, i.e., an imaginary dimension. Other transformations which stretch large distances would have the same effect. Another possible source of extra, uninterpretable dimensions might lie in individual differences between *Ss* in the manner of judging dissimilarity. Beyond the two big dimensions, a great variety of criteria can be used to judge additional distance between stimuli. If different *S*s use different criteria, the general result may well be "hash." As a final explanation or rationalization, we submit that it is perhaps too extravagant an expectation that minor dimensions should have a ready interpretation with such complex stimuli as facial expressions. At all events, the first two dimensions emerge strongly and clearly, embodying the important outcome of this study.

A persistent concern in the facial expressions area has been the shape of the stimulus surface (the "emotion solid"). Schlosberg (1952) designated the shape as roughly oval in two dimensions and (Schlosberg, 1954) as a tilted cone with an oval cross-section in three dimensions. Triandis and Lambert (1958) strove inconclusively to modify Schlosberg's formulation. One difficulty besetting these attempts has been the lack of calibration between dimensions.

Shape is a direct outcome of the multidimensional scaling operation. A plot of Dimensions I and II of the present study yields a stimulus boundary or envelope which is nearly an equilateral triangle. The "sleep" expression, Stimulus 56, anchors one corner of the triangle at the negative pole of Dimension II. The other two corners lie at the other pole of Dimension II such that the side joining them is parallel to the I axis; Stimulus 15, very pleasant surprise, is close to one corner and Stimuli 29 and 51, anger and fear, close to the other. The boundary could of course bulge were more stimuli added, with a coat-of-arms or pear-shaped contour as plausible possible alternatives. An oval or elliptical shape appears implausible, although one cannot have full confidence in this conclusion without a much more exhaustive sampling of stimuli.

A final word of clarification is in order. This study and all others like it can only achieve a representation of facial expressions *as interpreted by the viewer*. There is sometimes an implication in studies of this type that one might somehow identify the full range of possibilities for the "sender," i.e., discover the "true" dimensions of emotion. It is unwarranted to expect to complete this ambitious task by procedures involving viewers' judgments. Nuance is blurred or lost through misinterpretation, though this loss is not nearly so complete as was supposed in the earliest studies.

SUMMARY

Thirty women gave nine-point rating scale judgments of the dissimilarity of emotional expression of each of 78 pairs of facial poses formed from combinations of 13 diverse stimuli of the Lightfoot Series. The dissimilarity ratings were converted to interstimulus distances. Comparison of these distances with the distances one would expect on the basis of the three Schlosberg scales of facial expression yielded the following conclusions: (*a*) The pleasant-unpleasant scale explains 50% of the variance in the distance data; the pleasant-unpleasant (V_1) and tension-sleep (V_3) scales together account for almost 75% of the variance; the attention-rejection (V_2) adds little more to prediction. (*b*) If the V_2 scale is omitted, the V_1 and V_3 scales should be calibrated in approximately the ratio 5:4. (*c*) Standard multidimensional factoring techniques applied to the distance matrix yielded stimulus coordinates on five dimensions, with the major dimension essentially identical to Schlosberg's pleasant-unpleasant scale, the second dimension, a compromise between Schlosberg's other two scales, correlating extremely highly with both, and the remaining three dimensions not readily interpretable. (*d*) The shape of the configuration of stimuli in the space of the first two dimensions is approximately an equilateral triangle with a sleep stimulus at one corner and actively pleasant and unpleasant expressions at the other two corners. (*e*) If the Lightfoot Series has indeed captured the full range of potential expressions, then for usage a two-dimensional system would appear to be adequate: pleasant-unpleasant and tension-sleep, or perhaps a slight variant of the latter.

REFERENCES

DIEDERICH, G., MESSICK, S. J., & TUCKER, L. R. A general least squares solution for successive intervals. *Psychometrika*, 1957, 22, 159–173.

EKMAN, G. Dimensions of color vision. *J. Psychol.*, 1954, 38, 467–474.

ENGEN, T., & LEVY, N. Constant-sum judgments of facial expressions. *J. exp. Psychol.*, 1956, **51**, 396–398.

ENGEN, T.,, LEVY, N., & SCHLOSBERG, H. A new series of facial expressions. *Amer. Psychologist*, 1957, **12**, 264–266.

ENGEN, T., LEVY, N., & SCHLOSBERG, H. The dimensional analysis of a new series of facial expressions. *J. exp. Psychol.*, 1958, **55**, 454–458.

HELM, C., MESSICK, S. J., & TUCKER, L. R. Psychophysical law and scaling modes. Princeton: *Educ. Testing Serv. res. Bull.*, 1959, No. RB-59-1.

HULIN, W. S., & KATZ, D. The Frois-Wittman pictures of facial expressions. *J. exp. Psychol.*, 1935, **18**, 482–498.

MELLINGER, J. J. A comparison of multidimensional scaling and similarity analysis. *Amer. Psychologist*, 1958, **13**, 375. (Abstract)

MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21**, 1–16.

MESSICK, S. J., & ABELSON, R. P. Scaling and measurement theory. *Rev. educ. Res.*, 1957, **27**, 487–497.

NUMMENMAA, T., & KAURANNE, U. Dimensions of facial expression. *Acad. Paedag. Jyvaskylaensis res. Rep.*, 1958, No. 20.

OSGOOD, C. E. Fidelity and reliability. In H. Quaster (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955.

OSGOOD, C. E. Cognitive dynamics in human affairs. *Public opin. Quart.*, 1960, **24**, 341–365.

SCHLOSBERG, H. The description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1952, **44**, 229–237.

SCHLOSBERG, H. Three dimensions of emodimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

SHEPARD, R. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–346.

TORGERSON, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika*, 1952, **17**, 401–419.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

TRIANDIS, H. C., & LAMBERT, W. H. A restatement and test of Schlosberg's theory of emotion with two kinds of subjects from Greece. *J. abnorm. soc. Psychol.*, 1958, **56**, 321–328.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received March 25, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 555–562

SPONTANEOUS RECOVERY INTERVAL AS A FACTOR IN REACQUISITION OF T MAZE BEHAVIOR[1]

JOHN W. COTTON
University of California, Santa Barbara

GLEN D. JENSEN
University of New Mexico

AND DONALD J. LEWIS
Rutgers University

The present study attempts to show the effects of spontaneous recovery intervals ranging from 1 min. to 9 days upon reacquisition of both speed and correct turns in a T maze. This is a wider range of intervals than previously studied. Further, most available data are from other apparatuses. Data from the first reacquisition trial may be taken as an indication of spontaneous recovery, traditionally defined. Later reacquisition trials are studied as dependent variables of interest in their own right. A second purpose of this experiment is to provide definitive empirical curves of acquisition and extinction behavior in a T maze. For that reason experimental conditions are held constant for all animals during acquisition and extinction.

METHOD

Subjects.—The *Ss* were 100 naive male albino rats purchased from the Sprague-Dawley Company of Madison, Wisconsin. An additional 17 animals were discarded, 7 because of sickness or death, 8 because of apparatus failure or procedural error, and 2 because of failure to learn the running response. The *Ss* were from 55 to 80 days old at the beginning of experimentation, with a mean of 63.1 days. They were assigned at random to five groups, subject to the restrictions that no group have more than 1 more *S* completed at any time than any other group and that the total group size be 20 at the end of the experiment.

Apparatus.—The "narrow" T maze (2 in. inside width) used by Cotton, Lewis, and Metzger (1958) was employed. An automatic timing device measured the running time in .01-sec. units from the start box to a point $46\frac{1}{4}$ in. down the stem and a second timer began with the first and stopped when *S* reached the entrance to one goal box or the other. The readings of the second timer are not reported in this study. They were used as reported in the extinction procedure.

Procedure.—Six stages constituted the procedure. The period of deprivation and taming extended for 7 days in which each *S* was under 23-hr. food deprivation with 1 hr. of ad lib. feeding and a 5-min. period of handling each day. Each *S* was allowed to eat up to six 45-mg. lab chow pellets from a food cup on a table during every handling session. If *S* ate less than six during that session, the remainder were placed in the home cage immediately afterward.

A pretraining session on Day 8 consisted of 10 rewarded trials through the complete T maze for each *S* with a 15-sec. rest interval in the home cage between the end of one trial and the beginning of the next. Odd-numbered trials were free-choice trials; even-numbered trials were forced to the side opposite the preceding response. The deprivation period at the beginning of this and all subsequent sessions was 22 hr. except when stated otherwise below. One 45-mg. lab chow pellet was given on all rewarded trials following Day 7.

An acquisition session on Day 9 consisted of 30 noncorrection trials through the complete T maze. For each *S* the side less frequently chosen on free-choice trials of Day 8 was designated the correct side on Day 9. Turns to the correct side were then rewarded, and turns to the incorrect side were not rewarded. The rest intervals in this and succeeding periods were as on Day 8.

An extinction session on Day 10 consisted of 30 unrewarded trials through the complete T maze with the exception that, if *S* failed to stop the second timer within 2 min., he was removed from the maze. He then received a running time listing of 2 min. unless he had reached the microswitch stopping the first clock. Forty-three such removals were made during extinction but only 7 of these were before *S* stopped the first timer.

---

[1] This experiment was performed at Northwestern University under National Foundation Grant G-2672. The preparation of the article itself was supported by National Science Foundation Grant G-8706.

The spontaneous recovery period was a delay in the home cage between extinction and reacquisition. Members of the five groups were delayed 1 min., 10 min., 3 hr., 1 day, and 9 days, respectively. Groups 1D and 9D were placed on a 22-hr. food deprivation schedule until reacquisition began; Groups 1M and 10M had a deprivation period of approximately 23 hr. because the extinction session had added 1 hr. to their previous 22-hr. deprivation; similarly, Group 3H was under a deprivation of approximately 26 hr. This variation of 4 hr. in deprivation level may be expected to produce slightly faster running for Group 3H than otherwise. While the range from 22 to 26 hr. of food deprivation has not been explored, Yamaguchi's (1951) study of the range from 3 to 72 hr. suggests that the effect of variation from 22 to 26 hr. of deprivation is negligible.

A reacquisition period at the end of the spontaneous recovery period was identical for each *S* to the acquisition procedure.

RESULTS AND DISCUSSION

Comparability of groups.—Mean reciprocal running times were computed for each *S* for each stage of learning. Analyses of variance were conducted upon the means of the last 6 acquisition trials, the means of the 30 extinction trials, and the means of the last 6 extinction trials. None of these five analyses showed significant ($P < .05$) differences among groups, which is as expected since no differen-

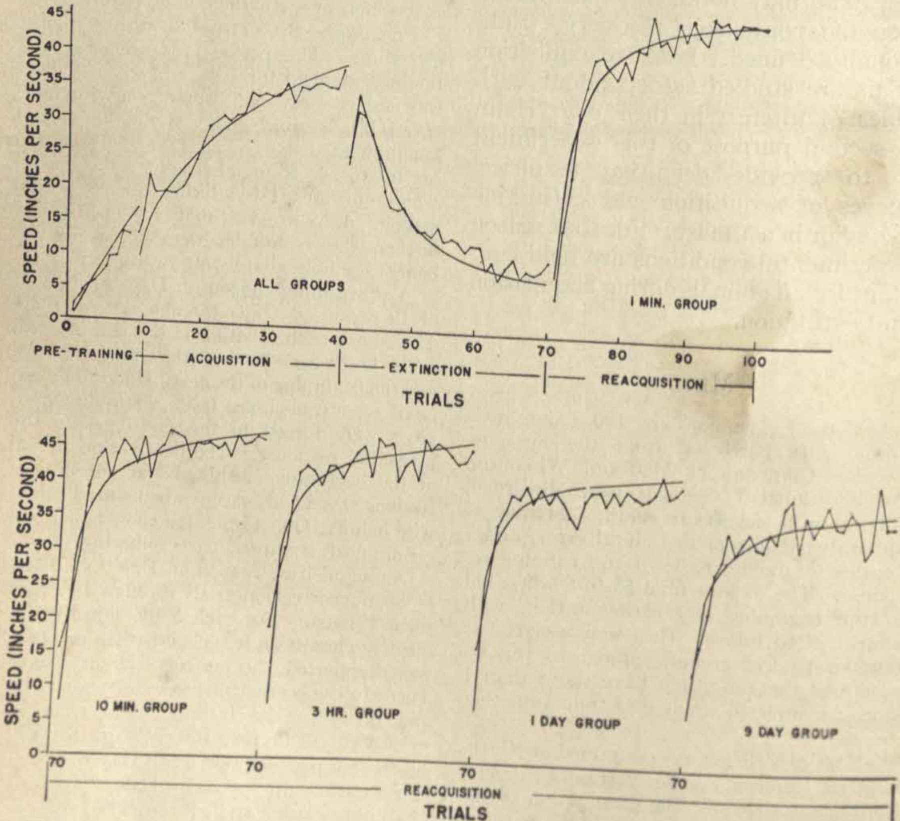


FIG. 1. A comparison of theoretical and empirical mean speeds, with data pooled for groups treated alike.

tial treatment was introduced until the spontaneous recovery period. The analysis of the last 6 acquisition trials gave a hint of initial differences among groups, with an F at the .07 level suggesting systematically slower running times by Group 9D.

Analyses of variance of numbers of correct responses for the same trial blocks just discussed also failed to show significant differences among groups. It is concluded that the group assignment procedure led to groups which may be considered comparable at the beginning of the retention period.

Effect of recovery interval.—The course of pretraining, acquisition, and extinction for all groups combined, and reacquisition of running speed for each group is shown in Fig. 1, together with fitted curves described later. Group 9D is worst in reacquisition, as it had been in pretraining and acquisition. Group 1M is also poor early in reacquisition but reaches the same level as Groups 10M and 3H, suggesting that minimal recovery over a 1-min. interval does not interfere with subsequent rewarded relearning.

TABLE 1

ANALYSES OF VARIANCE AND COVARIANCE OF RECIPROCAL RUNNING TIMES ON FIRST REACQUISITION TRIAL TOGETHER WITH OBSERVED AND ADJUSTED MEAN SPEEDS

Source	Variance Analysis			Covariance Analysis		
	df	*MS*	*F*	*df*	*MS*	*F*
Among groups	4	.193	3.06*	4	.168	2.87*
Within groups	95	.063	—	94	.058	—

Group	Observed Mean Speed (In/Sec)	Adjusted Mean Speed (In/Sec)
1 min.	9.71	10.13
10 min.	18.68	18.78
3 hr.	21.46	21.00
1 day	19.75	19.56
9 days	16.97	17.16

*$P < .05$.

The mean reciprocal running times for the first reacquisition trial were assessed by analyses of variance and covariance, using the mean reciprocal running time for each S on the last 6 extinction trials as the predictor variable. Table 1 summarizes these two analyses, showing that the five groups differ significantly on the first reacquisition trial regardless of whether or not covariance analysis is employed. Because the $r_{\text{within groups}}$ is only .28, the covariance analysis made little change in the F for group differences and no more than .01 difference of any adjusted mean speed from the observed mean. Table 1 also presents the observed and adjusted mean speeds (reciprocal running times multiplied by 46.25, the distance run while timed) for each group. The trend in each case is for substantial spontaneous recovery from Groups 1M to 10M, lesser recovery from Groups 10M to 3H, and then a decline in recovery for Groups 1D and 9D. Since the means for Groups 1D and 9D remain higher than that for Group 1M, these groups show a residual spontaneous recovery. The reduction in speeds for Groups 1D and 9D could indicate forgetting of the running habit. However, since a Tukey simultaneous interval test (Scheffé, 1959, pp. 73–75) shows no significant differences between groups except for the comparison of Groups 1M and 3H, it is more likely a random effect. Note that the four greatest delay groups show a mean increase of 70% or more above the observed mean speed of Group 1M.

Analyses of variance and covariance of mean speeds for all 30 reacquisition trials, using speeds from the last 6 extinction trials as the predictor variable, again both showed significant group differences. These analyses are not tabled here because of their similarity to the contents of Table 1. Both observed and adjusted mean speeds increase from Group 1M to Group 10M and decrease thereafter. No increment in speed compared with Group 1M appears beyond the 3-hr. condition. The mean for Group 9D is 5.26 in/sec below that of any other group. Analyses of variance and covariance omitting Group 9D yield moderate but nonsignificant *F*s ($P > .05$ in each case).

The mean percentages of correct responses for all groups combined in acquisition and extinction and for each group separately during reacquisition appear in Fig. 2, together with fitted curves to be described later. Percentages are not graphed for pretraining because even-numbered trials were always forced, and no differential reinforcement had yet been instituted. In early reacquisition trials Groups 1M and 9D appear inferior. No suggestion of group differences prior to the reacquisition period was found.

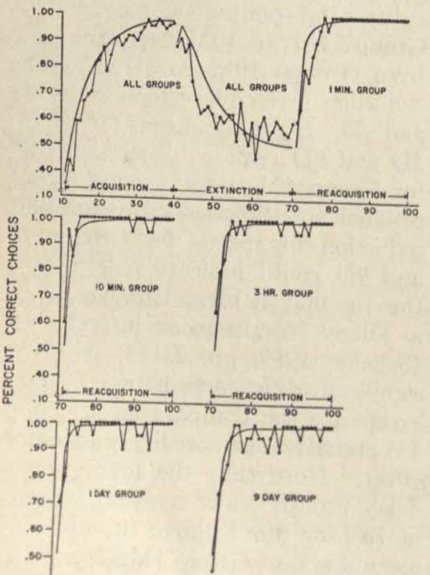


FIG. 2. A comparison of theoretical and empirical percentages of correct choices, with data pooled for groups treated alike.

Statistical analysis of turn data from the first reacquisition trial gives no indication of recovery interval effects. The mean percentage of correct responses for all groups combined is 60 on that trial, exactly the same as on the last extinction trial. The percentages of *S*s making the correct response on the first reacquisition trial are 60, 60, 65, 70, and 45 for the five groups arranged in order from 1M to 9D. The χ^2 obtained in a test for differences among these five percentages is 2.92, far below the 9.5 required for significance at the 5% level with $df = 4$. Similarly, a median test comparing the number of *S*s per group having more than the median number of correct responses in all 30 reacquisition trials yields a χ^2 of 7.55 ($P > .05$, $df = 4$). The trend of these numbers above the median is, however, suggestive: 11, 8, 8, 6, and 3, with the groups arranged in the same order as above.

This last indication that reacquisition of the correct turn may be retarded with increasing recovery time is bolstered by a tally of the number of *S*s having all perfect turns following their first correct choice during reacquisition: These numbers are 17, 14, 11, 10, and 6, with the groups arranged in the same order as before. A 5 × 2 homogeneity test using these frequencies and the corresponding numbers of *S*s not having all perfect trials following one reinforcement yields a χ^2 of 14.6 ($P < .01$, $df = 4$).

Curve-fits of the data.—The presence of responses for 100 *S*s per trial, with every *S* treated identically throughout pretraining, acquisition, and extinction, makes these data useful estimates of population trial by trial means. Accordingly the mean running speeds and percentages of correct choice have been fitted, using empirical procedures subject to the restriction that Hullian concepts be applied whenever possible and parsimonious.

The population mean running speeds (R_S) are hypothesized to conform to the following basic equation throughout the study:

$$R_S = A + KHR \qquad [1]$$

where A is the lowest speed obtained or to be expected in a given stage of the study, K is Hull's (1951) incentive motivation concept, altered according to Dufort and Kimble's (1956) suggestion to increase gradually during acquisition or following other increases in reinforcement magnitude and to decrease gradually during extinction or other decreases in reinforcement magnitude. As a consequence K mediates transition phases in studies like those of Crespi (1942) and mediates extinction and reacquisition as well. Inhibition constructs have been omitted because K handles the extinction phenomena of this study and because of the vagueness and inconsistency of Hull's inhibition postulates (Cotton, 1955). The quantity H is Hull's (1943) habit strength concept, subjected to Spence's (1956) contiguity assumption that H increases whenever the response in question is made, rather than only when it is overtly rewarded. The parameter R (for retention percentage) is an expression introduced for the first time by this article. It is intended to reflect the well-known phenomenon (Estes, 1955) that first-trial running speeds (a day after rewarded trials) are depressed compared with speeds at the end of the previous day. The value of R is presumed unity except on Trial 1 of such days.

In a T maze study we may properly think of two K values, one for each side. These will be called K_C, for the side correct during acquisition, and K_I, for the side incorrect during acquisition. It seems acceptable to assume that R_S will be controlled in part by the greater of the two Ks whereas R_P, the probability of the correct response, will be controlled in part by the difference $(K_C - K_I)$ between them. Thus K of Equation 1 becomes:

$$K = \max(K_C, K_I) \qquad [2]$$

where:

$$K_C = (K_O - A_P)10^{-j/N_S} \qquad [3a]$$

during pretraining or acquisition and:

$$K_C = K'(10^{-kN_S}) \qquad [3b]$$

during extinction (where K' denotes the K_C extrapolated from acquisition theory for the first trial of extinction but corrected for the shift in minimum speeds from pretraining to extinction; i.e.:

$$K' = \frac{K_C H + A_P - A_E}{H} \qquad [4]$$

with K_C and H being extrapolations from acquisition theory) and:

$$K_C = (K_O - A_R)(10^{-j/N_S}) \qquad [3c]$$

during reacquisition.

The equation for K_I has the same form as Equation 3a (except for different N_S values) during pretraining but, because it is never rewarded thereafter, takes the form of Equation 3b for the remainder of the experiment. Since there is no shift in minimum speeds from pretraining to acquisition, the analogue of Equation 4 is much simplified.

Previously undefined quantities in the equations above include a rate of growth constant j for acquisition of K, a rate of growth constant k for extinction of K, a maximum value K_O for K based on the magnitude of reinforcement used in the study, and N_S, the number of previous trials in a rewarded series or the number of previous trials in an unrewarded series. All but the last quantity are estimated wholly from the data. Specification of N_S values depended on several considerations: The "correct" side in the present study was the nonpreferred side during pretraining. Because only 20% of free choices during pretraining were to the later correct side and because, despite the fact that forced choices during pretraining equalized experience on both sides, a strong position preference existed at the beginning of acquisition, it will be assumed that each of the 10 pretraining trials added an increment of .2 to N_S for the correct side and .8 to N_S for the incorrect side, yielding values of 2 and 8, respectively, at the end of pretraining. During acquisition, N_S for the correct side increased by 1.0 per trial whereas N_S for the incorrect side started over at 0 and increased by 1.0 per trial because that side was being extinguished.

The two other terms of Equation 1 are:

$$H = 7 = B(10^{-iN}) \qquad [5]$$

where B is a curve-fitting constant, i is a curve-fitting constant showing the rate of growth of H, and N is the number of previous trials of any kind and:

$$R = 1 - C(10^{-mD}) \qquad [6]$$

on Trial 1 of a day following rewarded trials; =1 otherwise, where C is a curve-fitting constant, m is a curve-fitting growth constant, and D is the number of previous days of experimentation minus 1.

Figure 1, already shown, presents the theoretical running speed curves obtained

TABLE 2

CURVE-FITTING CONSTANTS USED WITH THE FIVE SPONTANEOUS RECOVERY INTERVALS

Groups	Reacquisition	
	K_O	j
1 min.	48.10	.618
10 min.	49.02	.596
3 hr.	48.33	.513
1 day	43.48	.430
9 days	40.14	.678

All Groups (Pretraining and Acquisition)

B	i	A_P	K_O	j	A_E	k	A_R	C	m
.923	.01725	5.09	47.73	.499	11.33	.0672	11.66	.418	.0561

Note.—A_R is a consequence of the extinction theory, rather than a parameter. It is presented to simplify verification of the theoretical values in Fig. 1.

with Equations 1 through 6, together with the empirical curves on which they were based. Table 2 shows the parameter values used in Fig. 3. All these values are based upon the data of this study, except for C and m which were obtained from unpublished curve-fits of first trial decrements in 9 successive days of 100% reinforcement training (Group A of Jensen & Cotton, 1960). Table 2 shows that K_O decreases systematically on either side of Group 10M, making 10 min. the point of peak recovery as expected from the statistical analyses for all 30 reacquisition trials. The j parameter increases on either side of Group 1D, indicating that rate of reacquisition is greatest for either short or long recovery intervals.

The curve-fitting of the turn data from this experiment is wholly a consequence of the theory presented for speeds, so that the only way in which it depends on turn data is in the previous assumptions about N_S during pretraining. The theory for turn data is consistent with Hull's (1943) Postulates 12 and 16 and (except for the lower limit of integration) with Spence's (1956, p. 203) equation. We assume that momentary effective reaction potential ($\dot{E}$) is normally distributed (except for truncation) with mean K_CH or K_IH depending on the choice in question and with variance σ^2. Hull's Postulate 16 tells us that the correct choice will be made whenever $\dot{E}_C > \dot{E}_I$. It is readily proved that, if we assume truncation of the $\dot{E}_C - \dot{E}_I$ distribution at ±2.5 from its mean,[2] the probability of a correct response will be 81/80 of the normal curve integral from $z = -2.5$ to:

$$\frac{(K_C - K_I)HR}{\sqrt{2}\sigma}$$

where:

$$z = \frac{\dot{E}_C - \dot{E}_I - (K_C - K_I)HR}{\sqrt{2}\sigma}$$

These values have been computed for every trial in this experiment, and were previously shown in Fig. 2, together with the obtained relative frequencies of correct choice. The quantity $\frac{1}{\sqrt{2}\sigma}$ is determined by the parameters of Equation 1 plus the assumption that the asymptotic value of R_P is 1.0. This requires that the upper limit of the normal curve integral just mentioned be 2.5 which in turn fixes $\frac{1}{\sqrt{2}\sigma}$ at some value for acquisition, with a new value for extinction and another new value for reacquisition because of shifts in asymptotic values of $(K_C - K_I)HR$.

Figures 1 and 2 indicate that relatively good approximations to trial by trial data can be made by the modified Hullian system suggested. This is not surprising in the case of Fig. 1 since the system of equations used was developed to describe the data therein. Since the theoretical curves of Fig. 2 were developed without any reference to turn behavior except for the number of correct responses made during pretraining, they offer more weighty support to the system.

---

[2] This assumption is *not* equivalent to assuming truncation of the $\dot{E}_C$ and $\dot{E}_I$ distributions and using the resulting $\dot{E}_C - \dot{E}_I$ distribution as Hull might have preferred. The resulting distribution would not be a truncated normal distribution as desired.

A noteworthy aspect of Fig. 1 and 2 is that, despite the fact that acquisition levels of speed and correct response were below their asymptotes, extrapolation of the acquisition curves across extinction to reacquisition (just as if rewarded trials had been given instead of extinction) leads to the final reacquisition levels obtained. This phenomenon has been called a "reinforcing effect of extinction" (Cotton, 1957; Jensen & Cotton, 1960) though alternative interpretations are available.

The most striking discrepancies between theoretical and obtained points in these figures occur for early trials of acquisition (running speeds only) and first trials of extinction with both response measures. The early acquisition data are irregular, even though based on 100 *S*s. Possibly an appeal to frustration drive theory (Amsel, 1958) would permit a better fit of these data. The problem of first-trial extinction data is compensated for by successful prediction of the second-trial increments (during extinction), for speed and for probability of a correct choice, a consequence of *R*'s increase from Trial 1 to Trial 2 of extinction. The discrepancies on Trial 1 of reacquisition suggest that a first-trial parameter varying with spontaneous recovery interval might well be added to the system. Such an addition will not be proposed until more data are available. In view of some indication earlier that only Group 9D differs significantly from the other groups over the 30 trials of reacquisition it is not even certain that parameter changes already made from group to group reflect effects of the independent variable more than of random sampling of *S*s for the five groups. The principal import of the curve-fitting, then, is to summarize trial-by-trial data in the four stages of the experiment rather than to show well-established effects of the spontaneous recovery interval.

Covariance structure of acquisition speeds.—The existence of reciprocal running times for 100 identically treated *S*s during acquisition and extinction also permits answering a statistical question relevant to many studies of animal learning. The analysis of variance for repeated measurements first proposed for the two-group case by Lindquist (1947) and then extended for many groups by Kogan (1948) has been proved by Geisser and Greenhouse (1958) to require the assumption of constant variances and constant covariances from trial to trial, though Geisser and Greenhouse do provide a conservative test for use when this assumption is not met. A test of whether this assumption, hereafter called the symmetric covariance assumption, holds is given by Box (1950). Since no evidence in the psychological literature regarding the acceptability of the assumption is known to us, such a test is reported here.

The 30 by 30 variance-covariance matrix for acquisition reciprocal running times[3] was obtained and tested to show whether the symmetric covariance assumption holds. Since a χ^2 of 864.65 with $df = 463$ was obtained, far beyond that required for significance at the .001 level (this χ^2 is equivalent to a z score of 11.17), it seems most unlikely that the assumption holds and hence it seems that the Lindquist-Kogan test is not legitimate for these data. Had such a test been deemed necessary for these data, the conservative procedure of Geisser and Greenhouse would have been necessary.

Fast and slow starting behavior.—Because of informal observations in the Northwestern laboratories of "jerky," "emotional" behavior during extinction, and related findings by Miller and Stevenson (1936), a feature of the present study was the recording of subjective judgments of "fast" or "slow" starts to the runs through the T maze. It was thought that extinction runs, though slower in over-all speed, might be marked by fast initial runs followed by refusal to enter the choice point or goal box or by fast retracings to the start box. A comparison of the percentages of fast starts during acquisition, extinction, and reacquisition[3] indicates that this prediction was in error. No statistical analysis was performed on these data, but all group percentages during extinction are lower than any group percentages in acquisition or reacquisition.

---

[3] A 3-page table presenting this matrix and a 1-page table presenting these percentages have been deposited with the American Documentation Institute. Order Document No. 7072 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

SUMMARY

One hundred male albino rats were given identical treatment during 10 pretraining trials with all responses rewarded, 30 acquisition trials with correct responses only rewarded, and 30 extinction trials with no responses rewarded, all in a single-unit T maze. Then randomly assigned groups of 20 *S*s received spontaneous recovery periods of 1 min., 10 min., 3 hr., 1 day, and 9 days, respectively, prior to a reacquisition session with all responses to the previously correct side rewarded.

Both observed and adjusted (by covariance analysis) mean speeds for the first reacquisition trial indicate systematic increases in running speed from 1 min. through 3 hr., followed by a nonsignificant decline (though not to the 1-min. level) through 9 days. Mean speed increases of 70% or more above the 1-min. condition were shown by the other four groups. Analysis of speeds for all 30 reacquisition trials combined gives similar results except that mean speeds begin to decline after 10 min. and no increment over the 1-min. condition appears in the 1-day or 9-day groups. The corresponding analyses of correct turn data do not reach statistical significance, but a χ^2 test shows that the number of *S*s having all correct turns following their first correct choice in reacquisition declines significantly with increases in the spontaneous recovery interval.

Neo-Hullian equations are presented and used for fitting mean running speeds and percentage of correct responses for each group on each trial.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BOX, G. E. P. Analysis of growth and wear curves. *Biometrics*, 1950, **6**, 362–389.

COTTON, J. W. On making predictions from Hull's theory. *Psychol. Rev.*, 1955, **62**, 303–314.

COTTON, J. W. The reinforcing effects of extinction. Paper read at Midwestern Psychological Association, 1957.

COTTON, J. W., LEWIS, D. J., & METZGER, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

DUFORT, R. H., & KIMBLE, G. A. Changes in response strength with changes in the amount of reinforcement. *J. exp. Psychol.*, 1956, **51**, 185–191.

ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

GEISSER, S., & GREENHOUSE, S. W. An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Ann. math. Statist.*, 1958, **29**, 885–891.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

JENSEN, G. D., & COTTON, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, **60**, 41–49.

KOGAN, L. S. Analysis of variance-repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.

LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend curves. *Psychometrika*, 1947, **12**, 65–78.

MILLER, N. E., & STEVENSON, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

YAMAGUCHI, H. G. Drive (*D*) as a function of hours of hunger (*h*). *J. exp. Psychol.*, 1951, **42**, 108–117.

(Received March 30, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 563–567

OPERANT CONDITIONING OF THE GSR[1]

R. L. FOWLER[2] AND H. D. KIMMEL

University of Florida

A controversy regarding the distinction between operant and respondent conditioning (Skinner, 1938) has been whether or not unelicited *autonomic* responses can be strengthened by the presentation of reinforcement after their emission, i.e., as operants. Konorski and Miller (1937) have asserted that autonomic reactions cannot be conditioned as operants, basing their conviction on the grounds that such responses are not instrumental in nature. Skinner (1938), while noting that children learn to cry "real" tears, tended to favor Schlosberg's (1937) notion that autonomic responses are conditioned classically and somatic responses instrumentally.

Kimble (1961) has defended the autonomic-classical and somatic-instrumental dichotomy by citing the unsuccessful attempts by Mowrer (1938) to condition the GSR instrumentally and by Skinner (1938) to condition the vasomotor reflex in operant fashion. Kimble's (1961) conclusion "Thus, for autonomically mediated behavior, the evidence points unequivocally to the conclusion that such responses can be modified by classical, but not instrumental, training methods," (p. 100) is perhaps the most "unequivocal" aspect of the controversy to date.

Recent evidence (Kimmel & Hill, 1960; Mandler, 1960; Razran, 1961) suggests that autonomic responses can be conditioned instrumentally. In the Kimmel and Hill (1960) experiment it was found that the rate of emission of unelicited GSRs was influenced by operant training methods, although the observed differences occurred *after* the reinforcement was terminated rather than during the reinforcement period. In the Russian study reported by Razran (1961) vasodilation (rather than vasoconstriction) became the predominant reaction to electric shock when it was made instrumental in shock termination.

In the present experiment some *S*s received reinforcement following the emission of GSRs while others received the reinforcing stimulus when not emitting GSRs. The reinforcing stimulus was a dim white light presented to *S*s who otherwise received no visual stimulation at all. In addition, the duration of the reinforcement period was varied. It was expected that groups which received reinforcement for twice the length of time would show a greater effect of such reinforcement. It was further expected that the differences during reinforcement, if any occurred, would tend to vanish during extinction.

METHOD

Subjects.—The *S*s were 40 students from introductory psychology classes at the University of Florida. They were assigned to four groups of 10 each, with one restriction on randomness. The first *S* to appear at the laboratory in a given experimental hour was randomly assigned to either a long or a short condition and the next *S* to the opposite duration condition so that 2 *S*s could be run per hour.

Instructions.—The *S*s were instructed to relax and to refrain from abrupt movements and deep or uneven breathing, but to remain alert and attentive.

---

[1] Based upon a Master of Arts thesis done at the University of Florida under the direction of the second author.

[2] Now at the University of Tennessee.

Apparatus.—A ¼-in. square light stimulus, produced by a 35-mm. projector located behind a rectangular panel approximately 2 ft. 8 in. in front of *S*'s eyes, was the reinforcing stimulus. Neutral filters were used to attenuate the stimulus. The average illuminance of the light (measured with a Macbeth illuminometer) was 0.36 ft-c and its duration, controlled by a Hunter interval timer, was 1.0 sec.

The GSR was picked up as a dc resistance change from the palm and back of the *S*'s right hand by ¾-in. zinc electrodes, covered with a few drops of saturated zinc sulfate solution, in lucite cups filled with saline electrode paste. It was amplified and then recorded on an Esterline-Angus ink-writing milliammeter with a paper speed of 3 in. per min. The current through *S* was 40 μa. A signal-magnet pen marked the stimulus on the GSR record.

Design.—The independent variables were the contingency or noncontingency of the reinforcement and the duration of the reinforcement period. Contingent lights were presented immediately after *S* emitted a GSR of a predetermined standard magnitude, i.e., at least one-half the average magnitude of that *S*'s unelicited responses during an initial 2-min. rest period. Noncontingent lights were presented only at times when no GSRs were being emitted. Each *S* who received noncontingent lights was matched with a previously run *S* who had received contingent lights, so that the noncontingent *S* received the same number of lights per minute of the reinforcement period as had been received by his contingent pair-mate. Half of the contingent and noncontingent reinforcement *S*s received 8 min. of reinforcement and the other half 16 min. of reinforcement. All *S*s received 4 min. of extinction immediately after the reinforcement period.

Procedure.—When *S* arrived at the laboratory, the palm and back of his right hand were cleaned with alcohol and the electrodes attached. The *S* was then seated in a cushioned chair in an air-conditioned, sound-insulated, electrically-shielded room, and the GSR leads were attached. The instructions were then read and the doors to the experimental room closed, placing *S* in total darkness. The equipment and *E* were located in an adjoining room.

After a brief period during which the GSR amplifier was balanced and its calibration checked, *E* began a 2-min. rest period, marking its beginning and end on the GSR record. The reinforcement criterion for each particular *S*'s response was determined at this time.

The reinforcement period was then begun. The *S*s in the contingent reinforcement groups received the light whenever *E* determined visually from the GSR record that a response met the reinforcement criterion. In making these judgments, *E* tended to be lenient, reinforcing all responses that were close to the criterion. Contingent lights were presented by *E* as quickly as possible after the recorder pen had reached the response magnitude criterion value.

The *S*s receiving noncontingent reinforcement received their stimuli spaced over 1-min. periods according to the response frequency of the contingent group *S* to which each was being matched. Thus, if a particular noncontingent group *S* was to receive four lights during a given minute, the stimuli were arbitrarily spaced out over the 1-min. period at times when no GSRs were being emitted.

After the reinforcement period all *S*s were given a 4-min. extinction period during which no stimuli were presented. Then the equipment was removed and the interview conducted.

Interview.—Each *S* was asked if he had the impression at any time during the experiment that the light was contingent upon "something" he might have done. If a negative reply was given, *E* then told *S* to assume that the light had been controlled by his behavior during the experimental session and, once again, asked if he could think of any possible connection between his behavior and the presence or absence of the light. Only the data of *S*s who completely denied any such correlation were used.[3] The *S*s were then thanked and asked not to discuss the experiment with anyone.

RESULTS

The basic dependent measure was the number of responses emitted per minute of reinforcement and extinction. A response was counted only if its magnitude was approximately equal to or greater than one-half of the average magnitude of the responses emitted by that *S* during the rest period. No response was counted if it appeared within 5 sec. after the presentation of a light stimulus (i.e., was elicited by the light). These criteria applied to all *S*s.

---

[3] Five *S*s, distributed unsystematically among the groups, were replaced for this reason or for failing to follow the instructions regarding movement.

The basic response frequency measure was first transformed to $\sqrt{x+1}$ to overcome the skewness of its frequency distribution. Then, the transformed measures for each minute of reinforcement and extinction were expressed as percentages of the average transformed measure of the 2 min. of rest just preceding reinforcement. This was done separately for each *S*. Figure 1 describes the changes in the measure of transformed relative response frequency for all *S*s during the period of reinforcement, averaged for 2-min. blocks to reduce variability.

Of primary importance was a comparison of the contingent and noncontingent groups during the reinforcement period. It was expected that the effect of reinforcement would be observable by comparing the relative response frequency curves of the contingent and noncontingent groups. As is indicated in Fig. 1, some separation of the two curves (each based on data of 20 *S*s) occurred during the first 8 min. of reinforcement, although this difference was not statistically significant ($F = 2.67$, $df = 1/266$) and tended to be obscured by a marked initial tendency in both groups toward reduced response frequency relative to the 2-min. rest period. It was further expected that the divergence of the response frequency curves of the contingent and noncontingent groups would tend to increase as a function of the length of the reinforcement period. Referring again to Fig. 1, it can be seen that even further separation of the two curves (each based on the data of 10 *S*s) occurred during the last 8 min. of reinforcement.

TABLE 1

MEANS AND SDs OF PERCENTAGES OF TRANSFORMED RESPONSE FREQUENCIES

$[100\sqrt{\text{Test}+1/\text{Rest}+1}]$

Minutes of Reinforcement	Reinforcement			
	Contingent[a]		Noncontingent[a]	
	Mean	*SD*	Mean	*SD*
Rest[b]	2.44	0.41	2.37	0.44
1	79.0	18.33	73.2	22.34
2	79.6	18.63	65.6	17.57
3	71.0	16.75	63.8	14.52
4	69.8	18.73	71.7	17.87
5	66.3	13.68	63.3	19.51
6	69.0	21.70	59.1	15.52
7	69.9	22.90	61.5	16.18
8	70.8	22.14	62.2	18.18
9	78.6	21.86	56.4	14.64
10	77.6	21.07	52.4	12.66
11	62.2	25.14	58.5	16.93
12	71.8	31.12	57.1	10.28
13	77.9	39.55	59.4	15.29
14	76.5	38.17	53.9	15.14
15	78.8	40.04	55.7	15.48
16	74.6	30.86	50.6	9.65

[a] $N = 20$ for first eight trials and $N = 10$ for second eight trials.

[b] Computed from 2 min. of rest before training, transformed to $\sqrt{X+1}$. These values were used as a baseline for the percentages in the following rows of the table.

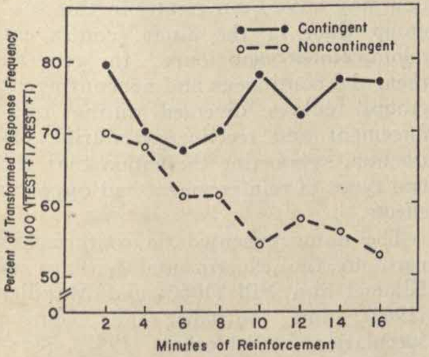


FIG. 1. Percentage of transformed response frequencies during 16 min. of reinforcement ($N = 20$ for each data point during Min. 1–8 and $N = 10$ during Min. 9–16.)

Table 1 gives the means and *SD*s of the transformed scores of the contingent and noncontingent groups,

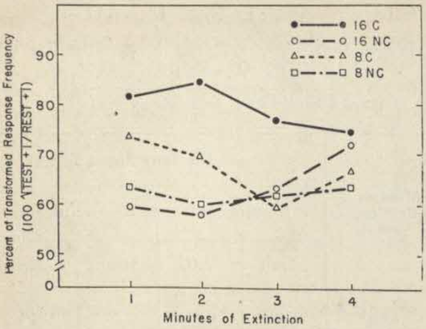


FIG. 2. Percentages of transformed response frequencies during 4 min. of extinction (for each data point, $N = 10$).

for the 16 min. of reinforcement. The analysis of variance for the entire 16 min., using only the measures obtained from the 20 *S*s who had received 16 min. of reinforcement, showed the Reinforcement Conditions × Minutes interaction to be significant ($F = 1.93$, $df = 15/270$). This interaction reflects the fact that the difference between the two groups grew throughout the reinforcement period, as had been hypothesized.

The data obtained during the extinction period provided another basis for evaluating the hypotheses regarding the effect of reinforcement. Figure 2 depicts the relative response frequency curves for each group separately for the 4-min. extinction period.

It was expected that the two contingent reinforcement group curves would show trends opposite to those of the noncontingent groups during extinction on the assumption that the different reinforcement procedures had opposite effects. Figure 2 suggests, in general, that the contingent and noncontingent curves showed opposite extinction trends, but the analysis of variance of these data provided only marginal grounds for rejecting the null hypothesis of no difference in group trends ($F = 2.25$, $df = 3/108$). The difference between the contingent and noncontingent groups' means was significant only for the first 1 min. of extinction ($t = 2.44$, $P < .025$).

DISCUSSION

The purpose of this experiment was to determine whether or not the GSR could be conditioned as an operant. The data support the conclusion that it can. Both acquisition and extinction response measures showed that the two types of reinforcement, i.e., contingent and noncontingent, led to differences in the rate of emission of unelicited GSRs.

Support was given also to the conclusion that the strength of conditioning depended upon the number of minutes that the unelicited response was reinforced. Finally, it was expected that the trends of the extinction curves for the two contingent subgroups would be opposite to those of the reinforcement curves for the same subgroups. The contingent group which received 8 min. of reinforcement did appear to show a general downward trend during extinction as opposed to an upward trend during the last 2 min. of reinforcement. The contingent group receiving 16 min. of reinforcement showed less systematic change during extinction, probably as a result of having received more reinforcement. That is, the resistance to extinction may have been greater in this subgroup than in the 8-min. contingent reinforcement subgroup. In general, then, the contingent and noncontingent groups' curves diverged during reinforcement and reconverged during extinction, supporting the notion that the two types of reinforcement had opposite effects.

The data presented do conform, in part, to the experimental findings of Kimmel and Hill (1960) and Mandler (1960), and contradict the negative speculations of Schlosberg (1937), Skinner (1938), and the direct assertions of Konorski and Miller (1937), and Kimble (1961).

To eliminate as much as possible,

the possibility that these results were due to "pseudo" operant conditioning (i.e., GSRs mediated by somatic movements which were actually conditioned) careful instructions were given to the *S* before the experiment began and a meticulous interview was conducted with *S* at its termination. All *S*s whose data were used not only stated that they had complied with the instructions, but also rejected the suggestion that their activities may have had any influence upon the occurrence of the reinforcing stimulus.

SUMMARY

This study attempted to determine whether or not the unelicited GSR can be conditioned instrumentally, using a visual reinforcer. The *S*s were 40 college students. All *S*s had 2 min. of initial rest, during which a base response frequency and a response magnitude criterion were determined. A reinforcement period followed the rest period. Two groups received 8 min. or 16 min. of reinforcement following each unelicited GSR. Two additional groups were matched to the first two groups in the number of visual stimuli presented per minute of reinforcement, but the stimuli were not given contingent upon the occurrence of unelicited GSRs. Finally, a 4-min. period of extinction was given. A significant interaction between reinforcement conditions and minutes of reinforcement was found in the 16-min. groups but not in the 8-min. groups. In general, the groups receiving contingent lights tended to respond at a higher rate than the noncontingent groups. During extinction, the relative response frequency trends for the contingent groups' curves tended to be opposite to those of the noncontingent groups (downward and upward, respectively), although this effect was not statistically significant. The results were interpreted as supporting the conclusion that the unelicited GSR can be conditioned instrumentally.

REFERENCES

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

KIMMEL, H. D., & HILL, F. A. Operant conditioning of the GSR. *Psychol. Rep.*, 1960, **7**, 555–562.

KONORSKI, J., & MILLER, S. On two types of conditioned reflex. *J. gen. Psychol.*, 1937, **16**, 264–272.

MANDLER, G. *Final report on Grant M-2442.* Washington: United States Public Health Service, 1960.

MOWRER, O. H. Preparatory set (expectancy): A determinant in motivation and learning. *Psychol. Rev.*, 1938, **45**, 62–91.

RAZRAN, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81–147.

SCHLOSBERG, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44**, 379–394.

SKINNER, B. F. *The behavior of organisms: An experimental analysis.* New York: Appleton-Century, 1938.

(Received April 15, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 568–574

AN ANALYSIS OF SET IN RELATION TO REACTION TIME

JACK BOTWINICK AND JOSEPH F. BRINLEY

National Institute of Mental Health

Statements or inferences about set (expectancy or readiness) are made from the relation between reaction time (RT) and the foreperiod or preparatory interval (PI) in a series. For example, Woodworth and Schlosberg (1954), in referring to an early study by Woodrow (1914), reported that "Maximum readiness . . . was not reached in much under 2 seconds and was not maintained much longer than 2 additional seconds" (p. 30). The reference was to RTs of 3 *S*s with a regular series.

In this statement by Woodworth and Schlosberg, set was described by two aspects—reaching and maintaining. RT set has also been described in terms of developing, organizing, mobilizing, and other aspects (e.g., Botwinick, Brinley, & Birren, 1957; Botwinick, Brinley, & Robbin, 1959; Huston, Shakow, & Riggs, 1937; Rosenthal, Lawlor, Zahn, & Shakow, 1960). In the present study the hypothesis that the various aspects of set are distinguishable, different in meaning, and independent was tested by principal component analyses (Hotelling, 1933). The analyses were based upon the intercorrelations of RTs for various PIs within various contexts. For example, if a "reaching" of set, as inferred from the RTs to short PIs, is independent from a "maintaining" of set, as inferred from RTs to long PIs, then clear components loaded by different PI durations might be expected. One component might be of short or relatively short PIs, and another component of long PIs.

Statements or inferences about set are also made from relations among RT-PI functions. For example, Huston, Shakow, and Riggs (1937), and Botwinick, Brinley, and Robbin (1959), compared RTs of regular and irregular series and discussed the relative slopes of the functions in terms of maintaining set. It is known that the value of RT in irregular series is related to the context in which the PI is embedded (e.g., Botwinick et al., 1957; Karlin, 1959; Klemmer, 1956), although its absolute duration may be a limiting factor (Botwinick, 1959, p. 749). Context even seems to make a difference in regular series, at least for schizophrenic *S*s, since PIs given in ascending order (increasing PI lengths) made for RTs that were different from those of PIs given in descending order (Zahn, Rosenthal, & Shakow, 1961). Thus the relations among parameters of RT-PI functions would appear to be contingent upon the specific contexts of the PI, as well as the absolute duration of the PI itself. In the present study, an attempt was made to assess the relative contribution of the absolute PI duration to that of its context. The same PIs in different series, varied in range and central tendency, were compared with respect to the associated RTs.

A third aspect regarding RT set involves sensory modality. While the particular sense modality is a factor in overall level of RT, it is not thought to be a determinant of set itself. Thus aspects of set are assumed to be independent of the sensory time courses. This assumption was tested as a hypothesis by comparing visual and auditory RTs in relation to the PIs of two types of series. In addition, these RTs were analyzed in relation to visual and auditory acuity of *S*.

METHOD

Subjects.—The *S*s were 32 men and 32 women volunteers. Mean age was 20.9 yr. (range 18 to 27 yr.), and mean education was 13.7 yr. (range 12 to 17 yr.). Twenty-seven of these *S*s (25 men and 2 women) were university students who came to the laboratory and were paid for their participation. The remainder of *S*s was quartered in an NIH ward for research purposes. All *S*s were in apparent good health with no known disease.

Procedure.—RT measurements were made with each *S* during three different sessions, each session lasting about 40 min. and taking place in most instances within 1 week, but never taking longer than a 2-week interval.

The total three-session period included 300 RT measurements under four conditions of auditory stimulation, and two conditions of visual stimulation. The four auditory conditions comprised three irregular series and a regular series. One regular and one irregular series comprised the visual series.

The visual and auditory regular series had PIs of 0.5, 1.0, 3.0, 6.0, and 15.0 sec. Half the *S*s were given the series in ascending order, and half in descending order. One auditory irregular series and one visual irregular series comprised these same five PI durations.

The remaining two auditory irregular series were of the following PIs: 0.5, 1.0, 1.5, 2.0, and 3.0 sec., and 3.0, 6.0, 9.0, 12.0, and 15.0 sec. All irregular series were arranged so that each PI followed every other PI an equal number of times.

All five PIs within each of the six series were presented 10 times. In the irregular series, in order to have each PI follow every other PI the same number of times, two additional measurements were made with the first and the twenty-seventh RT measurement (of the 52 in the series) not counted in the analysis.

Brief practice or warm-up RTs were given at the start of each series. Irregular series were preceded by two presentations of each of the five PIs in random order. Each PI of the regular series was presented twice with the 10 RT measurements for the PI commencing immediately after this.

Each of the three sessions comprised two series: one session was of the two visual series, another session was of the two comparable auditory series, and one session was of the auditory irregular series where the PIs ranged from 0.5 to 3 sec., and 3.0 to 15.0 sec. The order of the three sessions was systematically varied for the 64 *S*s, and the two series within the three sessions were counterbalanced as nearly as possible.

The *S* pressed down on a telegraph response key which started an automatically programed sequence. The finger press initiated a warning signal of 0.5 sec. duration, followed by the PI and then the stimulus. The stimulus was an 85-db., 1000-cycle tone presented through earphones, or the glow of a neon 40-w. lamp in front of *S*. The stimulus was terminated by the finger lift response. The warning signal was the glow of a 7½-w. frosted lamp in the case of the auditory stimulus, or a buzzer, approximately 64 db., in the case of the visual stimulus. Each *S* was encouraged to respond as quickly as possible.

Measures of auditory and visual acuity were made on each *S*. Air conduction audiograms were obtained with a Western Electric Audivox 7-B audiometer. Hearing loss in decibels of a 1000-cycle tone was measured for each ear in three ascending and three descending series. The median of the six series was used as the measure of hearing loss. Visual acuity measures were made of near and far vision (optical distances of 13 in. and 26 ft., respectively) with a Bausch and Lomb Orthorater. Standard test slides of binocular acuity were used.

RESULTS

Ten RT measurements were made for each PI within each series for each *S*. The median of each of these 10 measurements was computed and used in the analyses of the data.

PI in context.—It is clear from Fig. 1 that RT was related to absolute PI duration within a series, and to PI duration relative to the specific PI context. Thus for example, the four RTs with a PI of 3 sec. ranged in value from .172 to .213 sec. The four mean RTs were significantly different ($P < .01$) as determined by a variance analysis, and each of the four mean RTs were significantly different from each of the other three ($P < .02–.01$) as determined by t tests. In general, regular PIs made for quicker RTs than did irregular PIs although there was no reliable RT difference between the regular and irregular series with PIs of 15.0 sec. ($P > .05$). Reliable differences between RTs of regular and irregular series were found with the other four PIs of the series ($P < .01$).

The data in Fig. 1 also showed that statements regarding specific aspects of set based upon the slope of the functions need to be made in reference to the particular PIs in their context. Thus, the magnitude of the difference in RT between PIs of 1.0 and 0.5 sec. was contingent upon the range or central tendency of the series. Similarly, differences in RT between PIs of 6.0 and 3.0 sec. were contingent upon context. In each of these cases statistically significant differences were found between each RT difference tested against the other two ($P < .05–.01$). However, the absolute duration of PI, independent of context, appeared to be a limiting factor, at least with long intervals. With abscissa values longer than 6.0 sec., the top two curves in Fig. 1 followed very similar courses, and with PIs of 15.0 sec., the three RT values were almost identical.

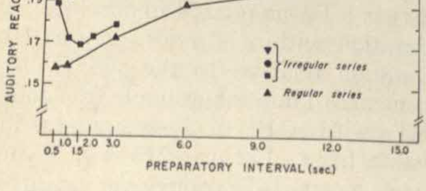


FIG. 1. Mean simple auditory RT of 64 *S*s as a function of the preparatory interval in four different series contexts.

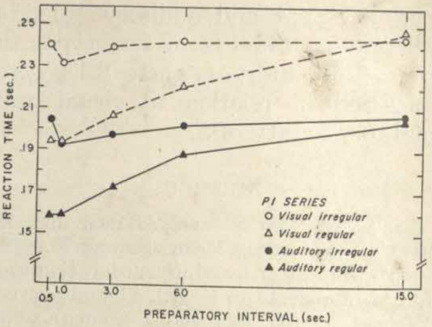


FIG. 2. Mean simple auditory and visual RT of 64 *S*s as a function of five preparatory intervals in regular and irregular series.

Sense modality.—It may be seen in Fig. 2 that RT was longer with the visual stimulus than with the auditory stimulus but that the increased latency was not a function of PI or its context. The data of auditory RT in this figure are the same as the corresponding data in Fig. 1. An analysis of variance was performed with the data of Fig. 2, but men and women *S*s were kept unpooled.[1] As shown in Table 1, the interactions between sense modality and PI, sense modality and type of series (i.e., regular and irregular), and sense modality, PI, and type of series, were not statistically significant ($P > .05$). However, visual RTs were longer than auditory RTs ($P < .01$).

In addition to the main effect of sense modality, the variance analysis included the main effects of PI, type of series, and also sex of *S*. It may be seen in Table 1 that all main effects and only one interaction (PI × Series) were significant ($P < .01$). This significant interaction indicated what may be seen in Fig. 2, viz., that the nature of the relation between RT and PI is different for the two types of series. Women *S*s were generally slower than men *S*s, different PIs made for different RTs, and the regular series made for generally quicker RTs than did the irregular series.

---

[1] We thank Seymour Geisser, Section on Mathematics and Theoretical Statistics, National Institute of Mental Health, for directing the variance analysis, and we thank Annie T. Randall, of the same section, for the computation involved.

In the next section the overall results of principal component analyses will be described. One of these analyses comprised the data of Fig. 2. As with the variance analysis, sense modality appeared independent of the PI. Separate components of visual and of auditory RT were not found.

Correlations of the visual and of the auditory acuity measures with the respective RTs ranged from low to zero. Visual and auditory acuity measures were also uncorrelated.

TABLE 1

ANALYSIS OF VARIANCE OF RTs TO VISUAL AND AUDITORY STIMULI

Source	*df*	*MS* (10000)	*F*
Sense mode (SM)	1	4545	544.98*
Prep. interval (PI)	4	420	98.73*
Series type (ST)	1	2020	185.50*
Sex (S)	1	274	21.21*
Individuals (I) within S	62	13	
SM × PI	4	1	0.55
SM × ST	1	8	1.90
SM × S	1	0	0.00
PI × ST	4	194	27.67*
PI × S	4	5	1.15
ST × S	1	1	0.11
SM × I	62	8	
PI × I	248	4	
ST × I	62	11	
SM × PI × ST	4	3	0.30
SM × PI × S	4	3	1.62
SM × ST × S	1	1	0.13
PI × ST × S	4	7	1.00
SM × PI × I	248	2	
SM × ST × I	62	4	
PI × ST × I	248	7	
SM × PI × ST × S	4	1	0.11
SM × PI × ST × I	248	11	
Total	1279		

* $P < .01$. All other *F* ratios not significant at .05 level.

TABLE 2

PRINCIPAL COMPONENTS I AND II OF SIX PREPARATORY INTERVAL (PI) SERIES: UNROTATED MATRIX

PI (Sec.)	Visual RT		Auditory Reaction Time			
	Reg.	Irreg.	Reg.	Irreg.	Irreg.	Irreg.
Components I						
0.5	65	81	53	87	89	
1.0	82	86	80	88	91	
1.5					89	
2.0					88	
3.0	85	88	91	87	91	89
6.0	90	85	85	85		93
9.0						93
12.0						90
15.0	84	89	80	88		94
%[a]	67	74	63	76	81	84
Components II						
0.5	67	48	78	32	39	
1.0	28	28	30	35	18	
1.5					−32	
2.0					−09	
3.0	−06	−23	−06	−16	−17	39
6.0	−31	−39	−30	−22		18
9.0						−18
12.0						−35
15.0	−40	−11	−42	−29		−04
%[a]	16	11	19	08	06	07

Note.—Loadings have been rounded from six places and decimals omitted.

[a] Percentage variance explained by component.

Components of set.—A Hotelling (1933) principal component analysis was performed on the RTs of the five PIs in each of the six PI series.[2] In each of the six component analyses, a first component was found which accounted for a large percentage of the RT variance.[3] In Table 2, it may be seen that the variance associated with these first components ranged from 63% to 84%, and all PIs within a series loaded positively and highly. The loadings within each of the four irregular seriēs were homogeneous with differences between them being only .03 and .05 in the auditory series, and .08 in the visual series. The loadings on the first component of the regular series were also homogeneous except for the shortest PIs (0.5 sec.). While the loadings of the 0.5-sec., regular PIs were high, they were substantially lower than those of the other four PIs in the series.

---

[2] The intercorrelations and the principal component analyses were directed by Donald F. Morrison of the Section on Mathematics and Theoretical Statistics, National Institute of Mental Health. IBM 650 digital computers were used for both the intercorrelations and the component analyses. We are grateful to Donald Morrison for this kind service.

[3] Copies of the correlation and principal component matrices will be furnished by the National Institute of Mental Health on request.

In Table 2 the second component for each of the six PI series may be seen. The variance accounted for by these components ranged from 6% to 11% in the four irregular series, and was 16% and 19% for the visual regular and auditory regular series, respectively. In each of the six series, the second component appeared to be related to PI duration, having high positive loadings for the shortest intervals and negative loadings for the longest intervals. Subsequent components, not included in Table 2, were uninterpretable, being less systematic in their bipolarity than the second components.

TABLE 3

PRINCIPAL COMPONENTS I AND II OF SIX PREPARATORY INTERVAL (PI) SERIES: ROTATED MATRIX

PI (Sec.)	Visual RT		Auditory Reaction Time			
	Reg.	Irreg.	Reg.	Irreg.	Irreg.	Irreg.
Components I						
0.5	44	66	27	76	78	
1.0	77	76	75	76	85	
1.5					94	
2.0					88	
3.0	87	91	90	88	92	78
6.0	93	92	86	88		88
9.0						94
12.0						95
15.0	87	89	86	92		93
%[a]	63	70	58	71	77	81
Components II						
0.5	89	67	97	53	58	
1.0	30	49	28	56	38	
1.5					−11	
2.0					11	
3.0	02	00	16	06	04	55
6.0	00	−16	08	00		35
9.0						00
12.0						−17
15.0	−01	12	−09	−06		14
%[a]	18	15	21	12	10	10

Note.—Decimals have been omitted from the loadings.

[a] Percentage variance explained by component.

The polarity associated with PI length in the second components suggested that with orthogonal rotations of axes, components could be obtained which reflected independent variance of the shortest PIs. Table 3 presents the orthogonal solutions by graphic rotations of the attempt to obtain such components. The rotated loadings involved the first two or three principal components of each of the six series.

The first component in each series was once again of general RT or RT set. The percentage of variance accounted for by each of these first components ranged from 58% to 81% with loadings of individual PI durations being relatively homogeneous, except again, for the 0.5-sec., regular PIs.

The second components of the six series were clearly identified with the short preparatory intervals. The rotations resulted in high loadings of short duration PIs and near zero loadings of longer ones. This was the case independent of regularity or irregularity of interval, and of the range and magnitude of PI durations within the series. Thus, both the absolute and relative shortness of the interval within a series was important in the organization of the short interval components, although the loadings of the shortest intervals were appreciably highest in the two regular series. The 0.5-sec., regular PIs of the short interval components accounted for more of the communality than these same PIs of general components. In all six series, however, the second components indicated that preparation during short intervals was distinguishable and independent of general RT.

In an attempt to discover short interval components common to more than one PI series, a principal component analysis was performed on RTs of the two regular series together.[3] Again, after rotation, a second component was found which was clearly identified as one of short interval. There were high loadings for the shortest PIs of both the visual and the auditory series, and negligible loadings for the longer intervals of both series. There was evidence of a similar nature for a short interval component common to the three irregular series which included the 0.5 sec. PIs.[3]

When both regular and irregular series were considered together, evidence did not appear for components loading on the shortest intervals. The RTs of all 30 PI conditions were intercorrelated and subjected to a principal component analysis. In addition, a component analysis was made of a 34 × 34 matrix in which the two visual and two auditory acuity measures were included. Component analyses were made also of RTs of the 20 PI conditions in Fig. 2 including the four acuity measures, and of the auditory RTs in Fig. 1 including the two auditory acuity measures.[3]

The first components of these four analyses were similar to each other, and to the first components of the six individual series. The first component in each of the four analyses accounted for approximately half the total variance, and all PIs loaded heavily on these components. Again, the loadings of the 0.5-sec., regular PIs were lower than the others. While the first components were easily identifiable, viz., general, the subsequent components were uninterpretable in their complexity. There was bipolarity that had no apparent order, there was no evidence of separate visual or auditory loadings, but there was some indication of regularity and irregularity of PI condition. Unlike the analyses of individual series, attempts to rotate the axes did not clarify the structure.

The 34 × 34 matrix of correlations indicated relations that were not apparent from the principal component analyses. Within the four irregular series, the RTs were intercorrelated highly among the PIs and were relatively homogeneous in magnitude. Within the two regular series, however, there were both high and low correlations, ordered with respect to the duration of PI. Thus, in the auditory regular series, the coefficient of correlation was .73 between RTs to the 15.0- and 6.0-sec. intervals. The coefficient was only .17 between 15.0- and 0.5-sec. intervals. The corresponding correlations among the visual regular RTs were .85 and .36, respectively.

DISCUSSION

The principal component analyses were performed to test the hypothesis that various aspects of RT set were distinguishable, different in meaning, and independent. In whatever PI series or group of PI series analyzed, a first component was found which accounted for an extensive amount of variance and which was clearly identifiable as a general RT or RT set component. These first components did not disappear on rotation. Short-interval components were also found, appearing as the second components in the rotated matrices of individual PI series, and in some of the series analyzed together. The short-interval components suggested at least one distinguishable aspect of RT set independent of overall set or RT level. The number of distinguishable aspects of RT set that were different in meaning, however, appeared to be limited in the present data. What other aspects were present, were ambiguous and too complex for identification. Long-interval components were not found, nor were components of meaningful combinations of long and short PI durations.

The meaning of the short-interval components appeared similar or general within regular series and within irregular series. In the two regular series together, and in the combination of irregular series involving the 0.5-sec. PIs, the shortest intervals within a series were grouped together. Thus, in regular series, the short interval components emphasized ability to become set for quick response within short time durations, and this ability was independent of the one for quick response when the interval indexing set was longer. The specific duration of the PI for this latter ability was of relatively little consequence. In irregular series, the short-interval components suggested ability to become set for quick response within PI durations that were short *relative* to other PIs in the series.

The short-interval components of the two regular series were larger in the proportion of variance explained than were those of the four irregular series, and perhaps were more important because of the very high loadings of the 0.5-sec., regular PIs. The meaning of the short interval may be different for the two types of series. Figures 1 and 2 indicate that RT was minimal with the 0.5-sec. PI in regular series but not in irregular series. When regular and irregular series were analyzed in one matrix, the 0.5-sec. PIs of the different series did not contribute to the variance of the components in the same way. In both types of series, however, RT variance associated with PI duration was found independent of overall RT variance.

The second aspect of RT set in this study involved the context of the PI. There is little doubt that RT was related to both the PI and its specific context, although context may be less important with long PI durations than with short ones. Thus, aspects of set, inferred from slope and other parameters of RT-PI functions, are related to PI contexts, and statements of RT set regarding populations, or comparisons of populations, must be made in reference to the specific conditions of measurement.

The remaining aspect of RT set investigated involved the role of sensation. Sense modality was a factor in overall level of RT, but not related to set as examined in this study.

SUMMARY

Principal component analyses were performed with reaction times (RTs) in relation to preparatory intervals (PIs) in six different series. Components of general RT level or RT set were found on which all PIs loaded substantially. In addition, short-interval components were found when matrices were rotated orthogonally. Long-interval components were not found, nor were other identifiable components. The results were discussed in relation to the hypothesis that aspects of RT set were distinguishable, different in meaning, and independent.

RTs were related to both the PI and its specific context. While RT level was different for auditory and visual stimulation, the difference was not related to PI or its context. Thus, set, as inferred from the relation between RT and PI, was independent of sense modality.

REFERENCES

BOTWINICK, J. Drives, expectancies, and emotions. In J. E. Birren (Ed.), *Handbook of aging and the individual: Psychological and biological aspects*. Chicago: Univer. Chicago Press, 1959. Pp. 739–768.

BOTWINICK, J., BRINLEY, J. F., & BIRREN, J. E. Set in relation to age. *J. Gerontol.*, 1957, **12**, 300–305.

BOTWINICK, J., BRINLEY, J. F., & ROBBIN, J. S. Maintaining set in relation to motivation and age. *Amer. J. Psychol.*, 1959, **72**, 585–588.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, **24**, 417–441, 498–520.

HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, **16**, 39–82.

KARLIN, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58**, 185–191.

KLEMMER, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

ROSENTHAL, D., LAWLOR, W. G., ZAHN, T. P., & SHAKOW, D. The relationship of some aspects of mental set to degree of schizophrenic disorganization. *J. Pers.*, 1960, **28**, 26–38.

WOODROW, H. The measurement of attention. *Psychol. Monogr.*, 1914, **17**(5, Whole No. 76).

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

ZAHN, T. P., ROSENTHAL, D., & SHAKOW, D. Reaction time in schizophrenic and normal subjects in relation to the sequence of series of regular preparatory intervals. *J. abnorm. soc. Psychol.*, 1961, **63**, 161–168.

(Received April 21, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 575–578

DONDERS' B- AND C-REACTIONS AND S-R COMPATIBILITY[1]

D. E. BROADBENT AND MARGARET GREGORY

Applied Psychology Research Unit, Cambridge, England

It was held by Donders (1868) that reaction time (RT) under b-conditions is longer than reaction time under c-conditions. That is, if a man is presented with two stimuli repeatedly in a random order, and if he has to make one reaction to one stimulus and a different reaction to the other stimulus, his RT is longer than it would have been if his instructions were to react only to one stimulus and to remain inactive to the other. This result has sometimes been regarded as demonstrating that the choice between two actions represents an extra process in addition to that of discriminating two stimuli. Donders himself suggested that the time taken by this extra process could be measured by b-c; but this involves the assumption that the two processes do not overlap in time, and this has usually been felt to be doubtful. The basic observation, however, is not normally questioned.

During the course of an investigation into verbal RTs, it was noticed that two groups of *S*s gave the same average time despite the fact that one was performing a b-reaction and one a c-reaction. Since the groups were small, this equality might have been due to chance factors; alternatively it was possible either that the classic experiments on the topic contain some loophole, or else that verbal RTs were abnormal in this respect. The classic experiments are not indeed very convincing to modern eyes. If for instance more than two stimuli are used, each receiving a separate response in the b-reaction but all except one ignored in the c-reaction, then there is a reduction in the number of stimuli discriminated as well as an abolition of choice between reactions. It is now of course realized (Crossman, 1953; Hick, 1952; Hyman, 1953) that such a reduction in the information transmitted through the nervous system may allow a reduction in RT quite apart from any other factor. However, there are cases in the literature in which only two stimuli were used and yet the usual difference between b- and c-reactions was found. Hick (1952), for example, in giving data for varying numbers of alternative reactions, includes the case where only one reaction is made. It appears from the experimental procedure that this case was in fact a c-reaction using the same two stimuli as the two-choice reaction, which is therefore the corresponding b-condition. The difference amounts to as much as 100 msec. in favor of c-reaction, so that it is quite clear and unambiguous. Thus the classic statements appear to be true even though the evidence which gave rise to them is not always convincing in the light of later knowledge.

There remains the possibility that verbal reactions are unusual. It may be recalled that Fitts and Seeger (1953) established that RT was faster if the type of response to be made was one which was particularly appropriate to the stimulus presented. That is, if a set of stimulus lights are arranged in a square, it is best for the response keys also to be arranged in a square, with each key corresponding to the light in the same position. The extent to which the responses were related to their stimuli in such a direct way was termed by Fitts and Seeger (1953) the "compatibility" of the S-R relationship. Now some recent evidence has suggested that an increase in the compatibility of a situation may abolish the usual tendency for RT to increase as the number of alternative reactions increases. Leonard (1959), for instance, has found that in responding to stimulation of a finger by pressing with that same finger, RT is largely independent of the number of fingers that must be discriminated. Pierce and Karlin (1957) found that the rate of reading, of English words seen clearly in

---

[1] Thanks are due to the British Medical Research Council for supporting this research; and to J. A. Leonard, not only for theoretical discussions, but for the loan and modification of an apparatus designed by him.

good illumination, was not affected by the size of the vocabulary in use, and it can perhaps be accepted that the most compatible response to a printed word is to say that same word. Mowbray and Rhoades (1959) have demonstrated for 1 *S* that, with very high degrees of practice, the difference between the time to choose one of four alternatives, and the time to choose one of two, will disappear even when the situation is the usual laboratory one of pressing keys in response to light signals.

Finally, Davis, Moray, and Treisman (1961) have shown that the time taken to repeat a word which is heard (a highly compatible task) does not depend upon the size of the vocabulary in use, and the unduly high speed of multiple choice reactions performed in this way has been observed independently by Chistovich, Aliakrinsky, and Abulian (1960).

While these workers have thus been indicating that only in relatively incompatible situations does RT increase with the number of alternative reactions, Rabbitt (1959) has shown that in such an incompatible situation it is the number of responses rather than the number of stimuli which primarily affects RT. If the same response is to be made to each of several different stimuli, it makes little difference how many stimuli are in fact involved so long as the number of responses is constant. Thus it is tempting to consider that the difference between the b- and c-reactions is of the same kind as that between two- and four-choice situations, both differences being due to the need to keep varying numbers of response tendencies in readiness. If this were so, the b- and c-conditions might well be identical in a highly compatible situation such as repeating heard words.

The present experiments were therefore designed to investigate the following questions: (*a*) Can the b- and c-reactions indeed be equally fast when the situation is one of repeating heard words? (*b*) Are the two conditions also equally fast in another highly compatible situation not involving the speech mechanism? (In fact, Leonard's tactile stimulation was employed.) (*c*) For both types of stimulus, does the traditional difference between b- and c-reactions reappear if the situation is made incompatible by changing the S-R relationship?

METHOD

Auditory reactions.—Stimuli were presented from a Ferrograph tape recorder through a loudspeaker mounted in a bass-reflex cabinet facing and touching *S* at knee height. The playback level was a comfortable listening one, i.e., it produced peak readings of about 75 db. on a Dawe sound-level meter. The responses were picked up by a Reslo ribbon microphone placed on top of the speaker cabinet and thus close to *S*'s mouth. Both stimulus and response were recorded on a second Ferrograph, and the time interval between them could afterwards be measured by locating them auditorily using small manual movements of the tape past the playback head: and measuring distance along the tape. This method of measurement possesses the advantage that the person making the measurement can ensure that the part of the response used is an informative one; and not merely a meaningless preliminary to speech as it may be with a voice key. On the other hand it is of course restricted to giving individual RTs to the nearest 10 msec. (about 2 mm. of tape).

The stimulus tape consisted of the words BID and DID arranged in random order, subject to the restriction that equal numbers of both stimuli occurred in each block of 30, and that no runs of more than 3 identical stimuli occurred. The interval between successive stimuli was 5 sec., and each was preceded by the word READY approximately 2 sec. earlier. There were four blocks, so that the whole tape totaled 120 stimuli, and the same tape was used for all conditions.

Tactile reactions.—The apparatus has previously been described by Leonard (1959). Briefly, *S* rested his fingers on relays, and a stimulus was given by applying 50-cps ac to the relay to produce a vibration. The response was pressure downwards on the relay. In the present experiment only the two index fingers were used, and the two corresponding relays were fitted each with a metal ring into which the finger was inserted. This prevented *S* from lifting one finger off the relay in the c-condition and so avoiding the stimulation of that finger altogether. RT was measured to the nearest 10 msec. by a Dekatron timer.

The sequence of stimuli used was identical to that for the verbal reactions, replacing BID by stimulation of one hand and DID by stimulation of the other.

Procedure.—Four groups of *S*s were tested,

TABLE 1

MEAN REACTION TIME (IN MSEC.) UNDER EACH CONDITION

Response	S-R Compatibility	*N*	b-Cond.		c-Cond.		*S*s Giving b > c
			Mean	*SD*	Mean	*SD*	
Verbal	Compatible	11	284	55.4	292	31.1	4
	Incompatible	6	444	31.2	354	57.7	6*
Tactile	Compatible	12	193	24.2	194	28.1	6
	Incompatible	6	300	25.3	255	19.5	6*

* $P < .05$ for b − c.

all Royal Navy enlisted men. No *S* served in more than one group. Each group was tested on 2 successive days, half the *S*s receiving the b-condition on Day 1 and the c-condition on Day 2, and half reversing this order. On each day a run of 120 stimuli was given, with a rest between each block of 30. The first 10 responses to the word BID (or with the right finger, as appropriate) were scored in each of the last two blocks. No difference in the trend of the results appeared between these blocks, and they have therefore been combined.

Group VC (Verbal Compatible) ($N = 11$) was instructed in the b-condition to respond to either word by saying the same word as was heard. In the c-condition they were to do this only for the word BID and were to ignore the word DID.

Group VI (Verbal Incompatible) ($N = 6$) was instructed in the b-condition to respond to the word BID by saying the word DECK, and to respond to the word DID by saying the word DIM. In the c-condition the word BID only was to receive the response DECK and the other stimulus was to be ignored.

Group TC (Tactile Compatible) ($N = 12$) was instructed in the b-condition to press down with the same index finger that was stimulated. In the c-condition they were to ignore stimulation of the left index finger and to respond only to the right finger.

Group TI (Tactile Incompatible) ($N = 6$) was instructed in the b-condition to press down the left index finger in response to stimulation of the right hand, and vice versa. In the c-condition they were to ignore stimulation of the right hand, and merely to respond with that hand to stimulation on the left side.

RESULTS AND DISCUSSION

Table 1 gives the mean RTs for each group in each condition. It is clear that both the groups working under compatible conditions give b-reactions that are slightly faster than their c-reactions. The difference is completely insignificant ($P > 0.5$ by a nonparametric test based on tau), and is in any event in the opposite direction to the traditional finding. The groups under incompatible conditions however give much faster c-reactions than they do b-reactions, and in each group every *S* shows this difference so that for each group $P < .05$ by sign test. Thus the traditional result seems to appear only with rather incompatible S-R relationships.

The RTs obtained are comparable with those found by Leonard (1959) in comparable conditions.

These results form part of the general body of data, cited in the introduction, which show that S-R compatibility interacts with the degree of choice present in a reaction time task. The theoretical implications of this empirical generalization have been discussed by Leonard (1960). It suggests a theoretical model of the type discussed mathematically by Stone (1960). Such models may be described as regarding the tendency to make any particular reaction as varying from moment to moment about a mean value, the variance reducing with the time for which the reaction is delayed. If the mean value is greater for the correct reaction than for erroneous ones, and if the reaction actually occurring is the one with the strongest tendency at the time chosen for response, RT will (for a constant probability of error) increase both with the number of erroneous reactions which are possible in the situation, and with the strength of the tendency to make any particular error. Such models include, as a special case, the familiar proportionality of RT to stimulus information: but they are of course related to decision theory rather than to the traditional informational analysis. In the present case, they would predict that the effect of adding an extra response to a task will depend upon the extent to which that response is one which competes strongly with the response already required in the task. The results are therefore consistent with a model of the decision theory type.

SUMMARY

It was hypothesized that the increase in RT under b- as opposed to c-conditions would vary according to the S-R compatibility of the particular situation used. Four groups of *S*s were therefore tested, two using verbal reactions to verbal stimuli, and two using key reactions to tactual stimulation of the fingers. One of the groups having each type of response used a compatible S-R relationship, i.e., they responded with a word identical to the word heard, or pressed a key using the actual finger stimulated. These groups showed no difference between the b- and c-conditions of Donders (1868). The other two groups used incompatible relationships, responding with a word other than the one heard, or with a finger other than the one stimulated. These groups showed the classic difference between b- and c-conditions. It can therefore be concluded that this difference is greatly reduced in conditions of high S-R compatibility.

REFERENCES

CHISTOVICH, A. A., ALIAKRINSKY, V. V., & ABULIAN, B. A. Time delays in speech repetition. *Prob. Psychol.*, 1960, **1**, 64–70.

CROSSMAN, E. R. F. W. Entropy and choice-time: The effect of frequency unbalance on choice response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

DAVIS, R., MORAY, N., & TREISMAN, A. Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, 1961, **13**, 78–89.

DONDERS, F. C. Die Schnelligkeit psychischer Processe. *Arch. Anat. Physiol., Leipz.*, 1868, 657–681.

FITTS, P. M., & SEEGER, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

LEONARD, J. A. Tactual choice reactions. Part 1. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

LEONARD, J. A. Choice reaction time experiments and information theory. *Proc. Lond. Symp. Inform. Theory*, 1960, **4**.

MOWBRAY, G. A., & RHOADES, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

PIERCE, J. R., & KARLIN, J. E. Reading rates and the information rate of a human channel. *Bell Sys. tech. J.*, 1957, **36**, 497–516.

RABBITT, P. M. A. Effects of independent variations in stimulus and response probability. *Nature*, 1959, **183**, 1212.

STONE, M. Models for reaction time. *Psychometrika*, 1960, **25**, 251–60.

(Received April 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 579–588

TASTE AND CONSUMMATORY ACTIVITY IN AMOUNT AND GRADIENT OF REINFORCEMENT FUNCTIONS[1]

FREDERICK A. KNARR AND GEORGE COLLIER

University of Missouri

In a prior runway study using sucrose reinforcements (Collier, Knarr, & Marx, 1961) it was concluded that variations in the total amount of consummatory activity are not necessary conditions for reinforcement. The present study is concerned with the question as to whether such variations are sufficient conditions. The taste of the reinforcing substance is varied conjointly with the amount of consummatory activity required to obtain the reinforcement.

One implication of the hypothesis that the behavior in the runway is, in part, chained by surrogates of the consummatory response is that the structure of the chain will be some function of the magnitude of the consummatory response. The present study examines this implication in terms of the variation in the gradient of running as a function of variations in the amount of consummatory activity as well as taste.

METHOD

Experiment 1

Subjects.—The *S*s were 53 male and 48 female albino rats, 90–115 days old at the start of the study. Twelve groups were matched on the basis of sex and predeprivation weights. One female from each group and a total of 5 males were discarded. Of the discarded *S*s, 3 *S*s failed to learn, 2 died, 2 were lost by procedural errors, and 10 were randomly discarded to equalize the number (7) of *S*s per group.

Apparatus.—A straight runway (Collier et al., 1961), 18 ft. long, $3\frac{1}{4}$ in. wide, and 5 in. high, enclosed in a light-tight, sound-deadened hull was used. The start box occupied the first $1\frac{1}{2}$ ft., the stem the following 14 ft., and the end box the remaining $2\frac{1}{2}$ ft. Guillotine-type doors, operated by motor driven screws, separated the start box and the end box from the stem. The floor consisted of nine pairs of 2 ft. long, $1\frac{1}{2}$ in. wide pieces of aluminum angle separated by $\frac{3}{16}$ in. side to side and by $\frac{1}{16}$ in. end to end. The times to traverse the last 6 in. beyond the exit of the start box and each of the next seven 2-ft. sections were recorded by means of electronic switches, activated by contact with the floor grids, operating .01-sec. Standard Electric timers. The time spent in retracing was included in the time to traverse the last section reached before retracing began. Thus eight times were obtained, a starting time and the times taken to traverse the seven successive sections preceding the end box. The start-box door opened 3 sec. after the start-box lid closed. The end-box door came down when *S* touched the ninth grid, 6 in. beyond the entrance to the end box.

The magazine consisted of a 50-ml. burette to which a glass drinking tube bent at an angle of 120° had been annealed. The tip of the drinking tube was presented $\frac{1}{2}$ in. behind a 1-in. diameter hole 2 in. above the floor in the rear wall of the end box. Two sizes of tips were used, 1 mm. and 4 mm. (internal diameter). Contact with the drinking tube activated an electronic switch which recorded the number of licks. The end box was illuminated by a 7-w. light located behind the hole for the tube.

The two concentrations (Percentage × Weight) of sucrose used were mixed from commercial sugar and distilled water 24 hr. before use.

---

[1] This investigation was supported in part by Research Grant M-3328 from the National Institutes of Health, Bethesda, Maryland. The computation in the analysis was performed by the University of Missouri IBM center and was supported by the Faculty Research Council. Experiment 1 is the first author's Master of Arts thesis. A brief report of Exp. 1 was given at the Midwestern Psychological Association meeting in St. Louis in 1960. Experiment 2 was performed by Earl Gates.

Procedure.—Three volumes per reinforcement (.5, 1.0, and 1.5 ml.), two concentrations (8 and 32%), and two tube size openings (1 and 4 mm.), were combined factorially.

The *S*s first received 10 days of deprivation accommodation, handling, and table training. A 22-hr. food and water deprivation schedule was used. Each *S* was handled for 3 min. per day followed by 5 min. of table top exploration. On the last 3 of the 10 days, *S*s were placed for 10 min. in the runway with the doors up and no reinforcement present.

Following the table training and handling *S*s were shifted to a 22-hr. food deprivation and an ad lib. water schedule. Ten days of magazine training were given, consisting of 3 days of drinking the appropriate concentration from the approprite burette in the home cage and then 7 days of drinking one trial per day in the end box.

Finally, each *S* was given 20 runway trials, 1 per day, on 20 consecutive days. The rat was removed from the end box immediately following his consumption of the appropriate amount of solution. Any *S* who failed to traverse the runway within 10 min. on 2 consecutive trials was discarded.

TABLE 1

Analysis of Variance of Starting Speeds, Exp. I

Source	*df*	*MS*	*F*
*S*s$_b$	83	.175	
Concentration (C)	1	6.72	73.16**
Tube size (T)	1	0	
Volume (V)	2	.032	.35
C × T	1	.011	.12
C × V	2	.220	2.39
T × V	2	.135	1.47
C×T×V	2	.189	2.06
Error$_b$	72	.092	
*S*s$_w$	756	.021	
Blocks (B)	1	1.969	37.11**
Trials (R)	4	.102	8.81**
B×R	4	.052	5.45**
C×B	1	1.765	33.26**
C×B×R	4	.03850	4.06**
C×T×B×R	4	.035	3.69**
C×T×V×B×R	8	.007	.71
Error$_w$	648	.015	
E$_1$	72	.053	
E$_2$	288	.012	
E$_3$	288	.009	

** $P \leq .01$.

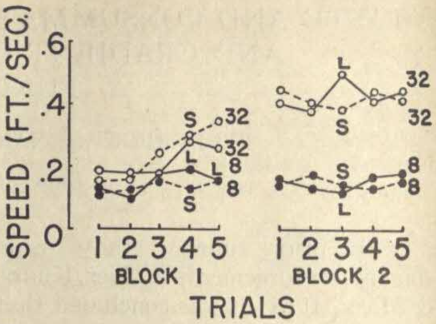


Fig. 1. Starting speed as a function of concentration, tube size, trial block, and trial.

Experiment 2

Subjects.—The *S*s were 15 female albino rats from the University of Missouri colony, 90–115 days old.

Apparatus.—An earlier version of the apparatus described above was used. The major difference was that the doors were solenoid operated. Noyes 45-mg. pellets were given as reinforcement in a glass coaster placed on the floor of the end box.

Procedure.—All *S*s received 7 days of deprivation accommodation to a 23-hr. schedule and 3 days of runway accommodation. The latter consisted of being placed 2 at a time for 6 min. in the runway in which Noyes pellets had been scattered. Following accommodation each *S* was given 35 trials, 1 trial per day.

Three magnitudes of reinforcement were used. The first group received 1 pellet and remained in the end box for approximately 2 sec. The second group received 8 pellets and remained in the end box for approximately 15 sec. The third group received 15 pellets and remained in the end box for approximately 30 sec. The end-box time represents the approximate time it took for *S* to consume the number of pellets offered as determined in preliminary trials on other rats. All *S*s were removed as soon as possible following consumption of their allotments. Fifteen pellets were used as the largest value since the preliminary experience showed that this was the maximum number of pellets that *S*s would consume without interruption.

RESULTS

Experiment 1

For purposes of analysis the first five and the last five trials were considered. All times were converted to speeds.

Table 1 presents an analysis of the starting speeds for concentration (C), tube size (T), volume per reinforcement (V), first and last block of trials (B), and trials (R). Figure 1 presents the highest order significant interaction, C × T × B × R. It is clear from Fig. 1 and the analysis that the 32% *S*s started faster; that they improved over trials and trial blocks while the 8% *S*s showed little if any improvement in starting speed over either trials or trial blocks; and that finally the large tube was superior for the 8% *S*s and the small tube for the 32% *S*s in the first block of trials and that there were no differences produced by tube size in the final block of trials. There was no evidence that volume per reinforcement had any significant effect on starting speed. Both the 8% and the 32% groups appeared to have reached a stable level of starting speed by 20 trials.

TABLE 2

ANALYSIS OF VARIANCE OF RUNNING SPEEDS, EXP. I

Source	*df*	*MS*	*F*
Ss_b	83	67.76	
Concentration (C)	1	2400.34	61.41**
Tube size (T)	1	35.55	.91
Volume (V)	2	12.45	.32
C × T	1	11.66	.30
C × V	2	47.96	1.23
T×V	2	71.94	1.84
C×T×V	2	48.56	1.24
Error_b	72	39.09	
Ss_w	5796	1.52	
Gradient (G)	6	60.59	61.20**
Blocks (B)	1	2854.55	444.63**
Trials (R)	4	106.36	82.45**
G×B	6	10.18	16.94**
G×R	24	2.65	5.76**
B×R	4	84.21	63.32**
C×G	6	11.21	11.32**
C×B	1	345.37	53.80**
C×R	4	13.77	10.67**
C×V×R	8	3.99	3.09**
C×G×B	6	6.47	10.77**
C×G×R	24	.71	1.53*
C×B×R	4	24.92	18.74**
G×B×R	24	2.23	1.64*
C×T×G×R	24	.72	1.55*
C×T×V×G ×B×R	48	.40	.29
Error_w	4968	.72	
E_1	432	.99	
E_2	72	6.42	
E_3	288	1.29	
E_4	432	.60	
E_5	1728	.46	
E_6	288	1.33	
E_7	1728	1.36	

* $P \leq .05$.
** $P \leqq .01$.

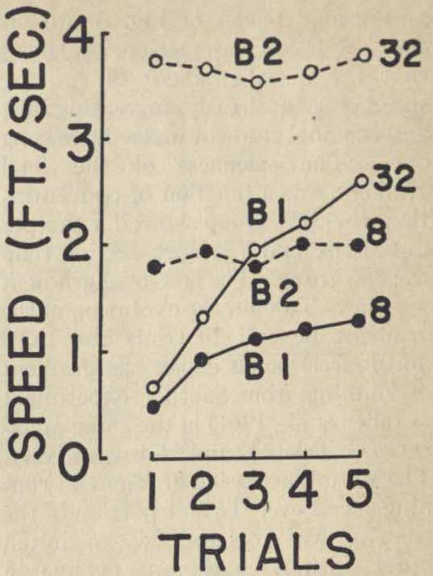


FIG. 2. Running speed as a function of concentration, trial block, and trial.

Table 2 presents an analysis of variance on running speed over C, T, V, B, R, and the sections of the runway, 2–8 (G). Figure 2 presents the C × B × R interaction. In the first block of trials the 32% group showed the most rapid learning, and in the second block of trials their performance was relatively stable while the 8% group was still improving. The final levels of performance of the 32% group were substantially greater than those of the 8% group. The first and second panels of Fig. 3 present the C × B × G interaction. It shows that the initial gradient of running speed was a slightly increasing one and the final gradient was a decreasing one. The steepness of the final gradient was a function of concentration, the 32% group showed a sharper initial rise and a steeper decline than the 8% group. The first-trial gradient was flat. To show the evolution of the gradient beyond 20 trials the third and fourth panels depict the gradient of running from another experiment (Collier et al., 1961) in the same apparatus on Trials 42 and 64, respectively. The initial increasing gradient in running speed over the first portion of the runway was replaced by an initial burst with a continuous decline in speed over the remainder of the runway. The remaining two interactions of interest were the C × V × R and C × T × G × R. The first resulted from the fact that for the 32% group, running speed ranked high to low for the 1.0, .5, to 1.5 ml. volumes per reinforcement, while for the 8% group the high to low running speed ranking was .5, 1.5, to 1.0. These differences increased slightly over trials. The differences in running speed resulting from the volume differences were relatively small. The second interaction, involving tube size, was the result of a small but persistent tendency on the early trials and the first portion of the gradient for the *small* tube size opening to produce faster running in the 32% group, in contrast to a larger, consistent tendency on the first portion of the gradient for the *large* tube size opening to produce faster running in the 8% group. These results are similar to those on starting times. Again, the differences produced by tube size were small in comparison with those produced by concentration. Tube size had its greatest effect at the low concentration.

Table 3 presents the average number of licks, the volume per lick, and the running and starting speed in each

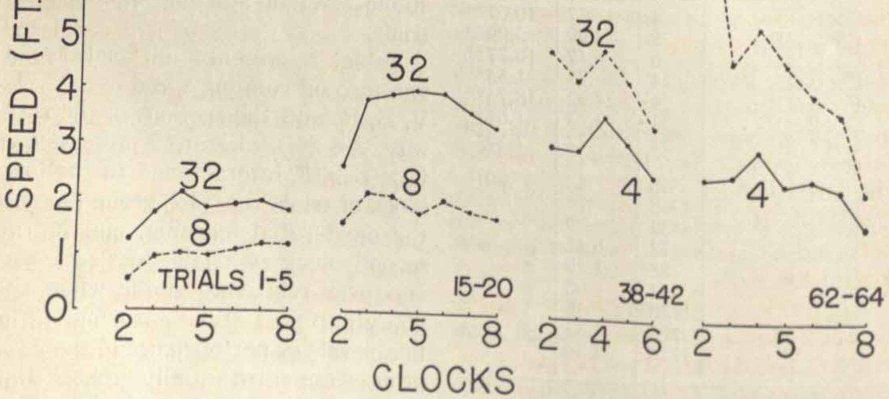


FIG. 3. Gradient of running as a function of concentration and trial block.

TABLE 3

LICKS, VOLUME PER LICK, STARTING, AND RUNNING SPEEDS AS A FUNCTION OF VOLUME PER REINFORCEMENT, CONCENTRATION, AND TUBE SIZE

Concentration %	Tube Size (Mm.)	Volume per Reinforcement (Ml.)	Licks	Volume per Lick (Ml.)	Starting Speed (Sec.)	Running Speed (Sec.)
8	1	.5	113.8	.0044	.13	1.58
		1.0	189.6	.0053	.14	1.85
		1.5	315.3	.0048	.18	1.81
8	4	.5	55.4	.0090	.11	1.96
		1.0	120.7	.0083	.19	2.22
		1.5	172.0	.0087	.14	2.03
32	1	.5	125.6	.0040	.39	3.39
		1.0	241.3	.0041	.43	3.60
		1.5	349.8	.0043	.43	3.90
32	4	.5	58.3	.0086	.48	4.21
		1.0	139.4	.0072	.29	3.08
		1.5	177.7	.0084	.50	3.85

group. An analysis of variance of the lick data for C, T, and V showed C, $P \leq .01$ ($F = 17.80$; $df = 1/72$); T, $P \leq .01$ ($F = 424.22$; $df = 2/72$); V, $P \leq .01$ ($F = 372.15$; $df = 2/72$); C$\times$T, $P \leq .05$ ($F = 5.64$; $df = 1/72$); and T $\times$ V, $P \leq .01$ ($F = 33.40$; $df = 2/72$) reaching significance. The 32% solutions required a significantly greater number of licks to get the offered volume which is reflected in the smaller volumes per lick at the higher concentration (Table 3). As would be expected, the small tube required a larger number of licks to obtain the required amount of solution. Tube size interacted with concentration; the difference between tube sizes was greatest at 32%. Volume and tube size also interacted, the largest volume per lick occurred at the large volume for the small tube size while the largest volume per lick occurred at the small volume for the large tube size.

The rank correlation between number of licks and starting speed was .35 ($P > .05$) for the 12 pairs of comparisons. The correlation between running speed and licks was .03 ($P > .05$). The correlation between starting and running speed was .72 ($P \leq .01$).

In summary, the rate of acquisition, the final level of performance, and the gradient of running were all functions of the concentration of reinforcement. By cross comparisons between the various volumes per reinforcement and tube size conditions it can be seen that these differences can not be attributed either to the number of licks, the volume consumed, or the duration of activity; i.e., the largest number of licks, the largest volume per reinforcement, or the longest time were not consistently associated with the fastest running. In fact, the fastest running time and the second fastest starting time are associated with next to the smallest number of licks.

The initial gradient of running for premagazine trained *S*s showed an initial rise and was relatively flat throughout the remaining distance. The final gradient showed an initial burst of speed with a progressive decline over the length of the runway.

TABLE 4

STARTING SPEED (FT/SEC) ON THE FIRST AND LAST BLOCK OF TRIALS

Number of Pellets	Blocks	
	First	Last
1	.166	.391
8	.214	.846
15	.230	.830

Experiment 2

Table 4 presents the average starting speeds for the three groups on the first and last block of the seven blocks of trials. An analysis of variance, Table 5, shows that both number of pellets and trial blocks were significant. The 15-pellet *S*s ran faster than the 8 and the 8 faster than the 1. All groups improved in starting speed over trials. There was no significant interaction.

Figure 4 shows the running speed over the various segments of the runway for the average of the first five trials and the average of the last five trials. An analysis of the data is presented in Table 6. The initial gradient is flat. The subsequent evolution is similar to that reported in Exp. 1, showing in its terminal phase, the fastest running in the first segment with a gradual decline in running speed in the direction of the end box. On the intermediate trials the fastest running occurred in the middle portion of the runway. The failure of N or any of its interactions to be significant appears to be the result of a single deviant *S* in the 1-pellet group who showed the fastest running and the steepest gradient of any *S*. When this *S* is dropped from the analysis and the group average in each block of trials at each segment is substituted, all of the interactions containing N prove significant at at least the .05 level. An examination of the individual curves for the 15-pellet group revealed that every *S* showed the gradient while only the deviant *S* in the 1-pellet group showed a consistent trend.

TABLE 5

ANALYSIS OF VARIANCE OF STARTING SPEEDS IN EXP. 2

Source	*df*	*MS*	*F*
Ss_b	14		
Number (N)	2	1.5069	3.88*
Error_b	12	.3880	
Ss_w	90		
Blocks (B)	6	.4629	12.25**
N × B	12	.0653	1.73
Error_w	72	.0378	

* $P \leq .05$.
** $P \leq .01$.

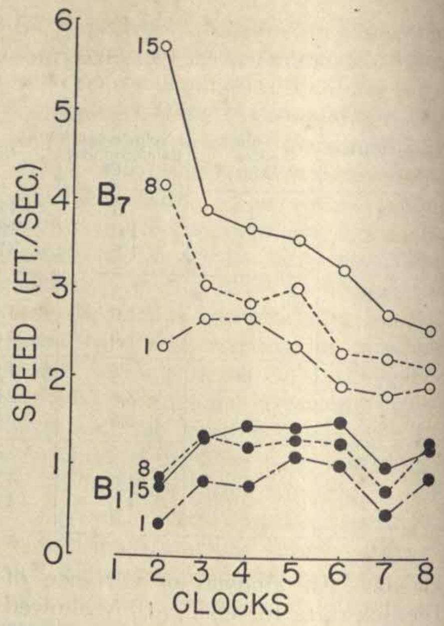


FIG. 4. Gradient of running as a function of number of pellets and trial block.

Figure 5 presents the speed of running for starting speed, and for the first and last segment of the runway plotted as a function of the number of pellets offered. The first trial curve is flat. In summary, the asymptotic gradient of running is negative and its steepness is a function of the magnitude of reinforcement. The speed of starting and running increases linearly with the number of pellets offered.

DISCUSSION

Amount of reinforcement.—It is clear from this and preceding studies (e.g., Collier et al., 1961; Kling, 1956; Stein, 1957) that variations in total number, average rate, or duration of licking, and volume or average volume consumed per second, are neither necessary nor even sufficient conditions for variation in acquisition or performance.

The studies of Stellar and Hill (1952), and Davis and Keehn (1959) show further that the momentary rates of licking or consumption are constant under a wide variety of conditions and, therefore, can not be considered determinants of the rate of acquisition or the level of performance.

An alternative account of the often found association between measures of consummatory responding and instrumental performance is that they are independent functions of some of the same variables. If we assume that the consummatory responses are respondents, whose strengths are functions of the intensive properties of the eliciting stimuli, and that one class of determinants of the amount of reinforcement function is the sensory consequences of the reinforcing substance (e.g., Collier & Myers, 1961), then such covariation would be expected. However, when one of these functions, for example amount of reinforcement as a function of volume, approaches an asymptote and the other does not, or when a given variable, for example tube size, affects one of these measures and not the other, then such covariation would not be expected. In those cases (e.g., Goodrich, 1960; Hellyer, 1953) where manipulation of the amount of consummatory activity appears to be a determinant of magnitude of reinforcement, it is possible that it gains this effect via the contribution of the consummatory responses to the sensory pool as well as the volume consumed per reinforcement.

TABLE 6

ANALYSIS OF VARIANCE OF RUNNING SPEEDS IN EXP. 2

Source	*df*	*MS*	*F*
Ss_b	14		
Number (N)	2	55.1173	1.59
Error_b	12	34.7511	
Ss_w	720		
Clocks (C)	6	9.7751	14.49**
Blocks (B)	6	45.3567	52.92**
B × C	36	1.3413	5.66**
N × B	12	1.3852	1.62
N × C	12	.8125	1.20
N × B × C	72	.2462	1.04
Error_w	576		
E_1	72	.6747	
E_2	72	.8571	
E_3	432	.2373	

** $P \leq .01$.

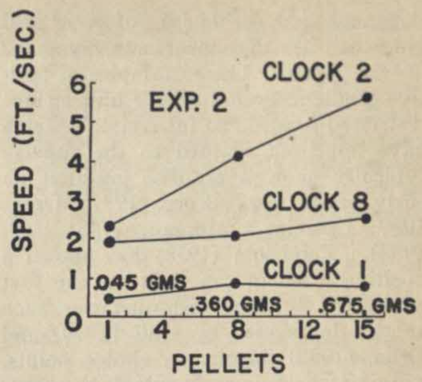


FIG. 5. Runway performance as a function of number of pellets.

Gradient.—The model of runway behavior that views it as a sequence of discrete responses pinned together by stimuli arising from the runway, from the deprivation operation, and from fragments of the consummatory response (e.g., Hull, 1932) implies a positive gradient of running, at least in the early acquisition trials. In spite of the textbook certainty and its appealing elegance, few examples of a "goal gradient" in the linear runway have been reported. These examples of positive gradients seem to occur under three restricted conditions: (*a*) Animals which have not been adapted to the runway typically show a positive gradient on early trials (e.g., Bruce, 1937; Drew, 1939; Epstein & Morgan, 1943; Hull, 1934). Pereboom (1958) has shown a positive gradient can exist on the first trial. (*b*) Runways with obstacles, such as the doors used by Hull (1934) and Drew (1939) for timing, choice points, running wheels (e.g., Meade & Newman, 1960), and delay doors (e.g., Wist, 1962) produce a positive gradient, at least in early trials. (*c*) Low deprivation (Bruce, 1937; Morgan & Fields, 1938) and small magnitudes of reinforcement (Crespi, 1942; and present study) produce a positive gradient. A further condition, which has not been systematically examined, producing a tendency toward a positive gradient is the use of open runways where there is a large amount of extraneous stimulation.

However, after extended training in a linear unobstructed runway, with a modicum of deprivation and reinforcement, the most characteristic gradient of speed of running is negative (e.g., Bruce, 1937; Collier et al., 1961; Crespi, 1942; Morgan & Fields, 1938; Weiss, 1960; and present study). The final gradient is a monotonically decreasing one with the fastest running occurring in the initial portion of the runway. The steepness of this final gradient is subject to such variables as amount of reinforcement (Crespi, 1942; Weiss, 1960; and present study), deprivation (Weiss, 1960), length of runway (Besch & Reynolds, 1958), and delay. The gradient, once obtained, shifts relatively rapidly with shifts in amount of reinforcement (Collier et al., 1961) or delay or position of delay (Wist, 1962). Thus it would seem that in its present form, at least, the goal gradient hypothesis can not account for the distribution of speeds of running in the runway.

An alternative account can be derived from the assumption that two classes of chaining of responses exist. The first class contains repetitive serial reflexes of a fixed, genetically determined topography such as licking, chewing, running, etc., in which the chaining stimuli are those produced by the preceding response. These sequences are not chained by external stimuli. They can be brought under the control of specific discriminative stimuli such as presentation of food, opening of a start-box door, etc. The second class contains responses of novel flexible evolutionary topography, such as right and left turns in a maze, sequential discrimination behavior, etc., which are chained by external stimuli. There exists also an anomolous middle class which has as members responses of the second class which are repetitive such as bar pressing and for which the chain shifts from external to internal stimulus control as learning proceeds. These two classes of chains have quite different properties, among which are the fact that the first acts as a unit, i.e., occurs in bursts of constant rate (e.g., Premack & Schaeffer, 1962; Stellar & Hill, 1952), the appropriate parameters being burst rate, burst duration, and burst latency; while for the second the individual response acts as a unit, the appropriate parameters being the probability and/or latency of occurrence. It is only for this latter class of chaining that the classical gradient of reinforcement would be expected.

If now runway behavior is considered as consisting of three independent classes of behavior, starting, running, and stopping, an account of the relations obtained can be given. Starting is a response elicited by an external stimulus and is highly sensitive to consequences, reinforcement, delay, etc., of the trip down the runway. Running is an internally chained response which occurs at a constant rate in bursts of varying duration and intervals between bursts. Differences in average running speed are the result of variations in these parameters of running. Premack and Schaeffer (1962) have further suggested that discrete burst rates exist as a function of the rat's gait. Thus most of the running in a linear runway is not under the control of an external sequence of stimuli but rather it consists of an internally chained sequence of responses of which only the initial member may be sensitive to the conditions of reinforcement, etc. The degree to which responses elicited by the nondiscriminative stimuli in the runway compete with starting and running appear to be a function of intra- and intertrial adaptation (e.g., Hull, 1934), amount of reinforcement (e.g., Armus, 1959; Pereboom & Crawford, 1958; and present study), and deprivation (e.g., Cotton, 1953). Burst duration and interval between bursts are functions of the parameters of reinforcement, deprivation, etc. The occurrence of a gradient of running may be attributed to the distribution of burst rates and burst durations. The probability of a fast burst is greatest in the initial portion of the runway since its initial member is starting, which is most sensitive to the conditions of reinforcement. A positive gradient of running is associated with a slow start however produced and a negative gradient with a fast start however produced. Such a relation can be observed in the individual trials where 72% of the block of last five trials for the 5 *S*s in the 15-pellet group of Exp. 2 were monotonic, that is had their fastest run in the first segment. Ninety-two percent had their fastest run in either the first or second segment. It seems likely that the shape of the gradient is the result of phenomena occurring in the start box rather than in the end box or in the rat's mouth.

SUMMARY

In two runway experiments the effects of jointly manipulating concentration of sucrose, volume per reinforcement, and amount of consummatory activity on starting and running speed were examined. Rate of acquisition and the terminal level of responding proved to be functions mainly of concentration and of number of pellets offered. The relatively minor effects of tube size occurred in the early training trials and those of volume per reinforcement in the terminal trials.

It was concluded that variations in consummatory activity are neither necessary nor even sufficient conditions for producing variations in acquisition and performance.

The implications of the fact that the terminal gradient of running was a negative one whose steepness was a function of concentration and amount were explored.

REFERENCES

Armus, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

Besch, N. F., & Reynolds, W. F. Alley length and time of food deprivation in instrumental reward learning. *J. exp. Psychol.*, 1958, **56**, 448–452.

Bruce, R. H. An experimental investigation of the thirst drive in rats with especial reference to the goal gradient hypothesis. *J. gen. Psychol.*, 1937, **17**, 49–62.

Collier, G., Knarr, F., & Marx, M. Some relations between the intensive properties of the consummatory response and reinforcement. *J. exp. Psychol.*, 1961, **62**, 484–495.

Collier, G., & Myers, L. The loci of reinforcement. *J. exp. Psychol.*, 1961, **61**, 57–66.

Cotton, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

Davis, J. D., & Keehn, J. D. Magnitude of reinforcement and consummatory behavior. *Science*, 1959, **130**, 269–271.

Drew, G. C. The speed of locomotion gradient and its relation to the goal gradient. *J. comp. Psychol.*, 1939, **27**, 333–372.

Epstein, M. A., & Morgan, C. T. Cortical localization of symbolic processes in the rat: III. Impairment of anticipatory functions in prefrontal lobectomy in rats. *J. exp. Psychol.*, 1943, **32**, 453–463.

Goodrich, K. P. Running speed and drinking rate as functions of sucrose concentration and amount of consummatory activity. *J. comp. physiol. Psychol.*, 1960, **53**, 245–250.

Hellyer, S. Duration of the consummatory response as a variable in amount of reinforcement studies. Unpublished doctoral dissertation, Indiana University, 1953.

HULL, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, **31**, 25–43.

HULL, C. L. The rat's speed of locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, **17**, 393–422.

KLING, J. W. Speed of running as a function of goal-box behavior. *J. comp. physiol. Psychol.*, 1956, **49**, 474–476.

MEADE, R. D., & NEWMAN, P. J. Locomotion gradients in the activity wheel. *J. comp. physiol. Psychol.*, 1960, **53**, 381–384.

MORGAN, C. T., & FIELDS, P. E. The effect of variable preliminary feeding upon the rat's speed-of-locomotion. *J. comp. Psychol.*, 1938, **26**, 331–348.

PEREBOOM, A. C. A goalless gradient. *J. exp. Psychol.*, 1958, **55**, 31–33.

PEREBOOM, A. C., & CRAWFORD, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56**, 82–85.

PREMACK, D., & SCHAEFFER, R. W. Distributional properties of operant-level locomotion in the rat. *J. exp. Anal. Behav.*, 1962, **5**, 89–95.

STEIN, L. The classical conditioning of the consummatory response as a determinant of instrumental performance. *J. comp. physiol. Psychol.*, 1957, **50**, 269–278.

STELLAR, E., & HILL, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45**, 96–102.

WEISS, R. F. Deprivation and reward magnitude effects on speed throughout the goal gradient. *J. exp. Psychol.*, 1960, **60**, 384–390.

WIST, E. R. Amount, delay, and position of delay of reinforcement as parameters of runway performance. *J. exp. Psychol.*, 1962, **63**, 160–166.

(Received April 28, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 589–592

ACTIVITY DURING DELAY OF REINFORCEMENT IN HUMAN LEARNING

R. A. CHAMPION AND D. A. McBRIDE

University of Sydney

Compared with immediate reward, delayed reward leads to inferior performance on the part of subhuman animals in learning situations. A number of attempts have been made to explain this phenomenon, the most recent being that of Spence (1956), who has supposed that a delay period allows the stimulus components to become attached to a variety of responses which compete with the response being learned, so that the longer the delay the greater the number of competing responses and the poorer the performance. This hypothesis was supported in a study by Carlton (1954) in which the effect of delay of food reward was significantly reduced in rats when competing responses were kept to a minimum by confinement during the delay period. On the other hand, a series of experiments on the effect of delay of knowledge of results in human learning (e.g., Bilodeau & Ryan, 1960) has revealed no comparable interference with performance, and some resolution of this apparent conflict in results is required. The demand which led to the following study has been explicitly stated by Noble and Alcock (1958) as an examination of human learning over a range of delay intervals both with and without disrupting interpolated activity. The present experiment was accordingly designed to serve two purposes: (*a*) a further test of Spence's hypothesis, and (*b*) an exploration of the possible application of the hypothesis to human performance.

With regard to the first of these purposes, Carlton had to be content with the assumption that the unconfined rats would be active during delay and in fact observed that some such animals were not active. In order to achieve a higher degree of control over activity during the delay interval, rats could be presented with a stimulus evoking a previously learned response or human *S*s could be made active under instruction; the latter alternative was chosen in the present case. This choice having been made, care was taken that Spence's hypothesis might be meaningfully applied to human *S*s, in keeping with the second purpose of the experiment. The effects of delay of reward in lower animals have mostly been obtained with latency or speed as a measure, applied to an act already high in the response hierarchy (such as lever pressing). The human *S*s were therefore simply required to give the response items for very easy paired associates and no measure of performance was taken unless the response was correct. A verbal response was preferred to a motor act because it allowed the presentation of the correct response item to be used as reinforcement. It should be noted that the term "reinforcement" is being used here in a purely empirical sense (Champion, 1960), for even performance with easy paired associates presumably changes little unless the response is followed by the correct item.

METHOD

Subjects.—There were 80 *S*s, recruited mainly from undergraduate classes at the

University of Sydney. The only requirements for selection were that *S* be free from speech defect and naive as to the purpose of the experiment.

Apparatus.—The material to be learned consisted of five paired associates as follows: NEEDLE-THREAD, SWEET-SOUR, TABLE-CHAIR, SHEEP-ANIMAL, STEM-FLOWER. These were made up on the basis of the Kent-Rosanoff study of free association (Rosanoff, 1938) such that the response in each case was the most frequent association and no stimulus had been given as a response to any of the other stimuli. The items were presented in a Ranschburg apparatus and constant exposure periods were provided by electronic timers. The latency of *S*'s response to each stimulus word was measured with an electronic chronoscope which was automatically started with the presentation of the stimulus and stopped by *E* as soon as *S* responded.

Two groups of *S*s were required to read words aloud during the period between their response and the appearance of the response word, and the interpolated words were also common associations with the stimulus word in the Kent-Rosanoff study. For example, the words to be read out after the response to the stimulus NEEDLE were: SHARP, PIN, SEW, STEEL, EYE, INSTRUMENT, POINT, THIMBLE, SEWING. These items were typewritten in groups according to the stimulus word on stiff 5 × 2½ in. cards hinged in an aluminum case, with tabs to facilitate the handling of the cards, and were placed at *S*'s right hand; the top two cards were blank. The paired associates and the corresponding sets of words to be read aloud were presented in three different orders, and the sequence of the latter words for any one pair was varied in each order.

Procedure.—Two periods of delay of reinforcement were used, 2 sec. and 5 sec.; half the *S*s on each delay period were required to read words aloud during delay whereas the remaining *S*s were instructed to fixate the slot in the exposure apparatus. This factorial design gave four groups designated henceforth as follows: (*a*) Group 2N: 2-sec. delay of reinforcement with no activity during delay, (*b*) Group 2A: 2-sec. delay with activity, (*c*) Group 5N: 5-sec. delay with no activity, and (*d*) Group 5A: 5-sec. delay with activity. The *S*s were allocated to these groups in successive cycles according to the order of their appearance at the laboratory, giving 20 *S*s in each group.

Each presentation began with *E* giving a verbal ready signal, followed about 1 sec. later by the activation of the exposure apparatus to present the stimulus word and to start the chronoscope. Immediately as *S* responded *E* simultaneously stopped the chronoscope, reactivated the exposure apparatus to present a blank to *S*, and started the timer controlling the delay-of-reinforcement interval. After 2 sec. or 5 sec. the exposure apparatus was automatically activated once more so as to present the response word for 1 sec., followed automatically by another blank. One run through the list of five items was regarded as a trial, and training was continued for 15 trials (75 presentations). The interval between presentations averaged approximately 20 sec.

Instructions.—The *S*s in Groups 2A and 5A were instructed as follows:

> This experiment is concerned with the learning of certain words as responses to other words. When a word appears in the slot in front of you, say the first word you think of. A few seconds after you give your reply a second word will appear and this is the response you are to learn. At first you will simply be guessing but the task is easy and you will learn it quite quickly. After you give your response, but before the correct response appears, you are to flip down the blank card thus (at this point the first blank card was pulled down as a demonstration) and read aloud the words which are typewritten on the card behind it. You are to pull down a card and start reading immediately after you make your response each time. Read these words aloud at the rate of about one per second. Stop reading as soon as the correct response word clicks into the slot. If you should finish reading the words on a card before the correct word has appeared, start reading at the beginning of the same card again. Remember that although the correct responses are easy to learn it is the speed with which you give your response each time that counts. I will say "now" about 1 sec. before each word to which you are to respond appears in the slot. Have you any questions?

After 15 responses had been given, so that all the word lists had been read from the cards, the *S*s in Groups 2A and 5A were instructed to push all the cards except the first blank back into their original position and to continue to do this each time all the cards had been read. The *S*s in Groups 2N and 5N were given the same general instructions as the other *S*s except that all references to reading words aloud were eliminated and they were instead asked to watch the exposure slot carefully throughout each trial, especially when waiting for the correct response word to appear.

RESULTS

Each raw latency score in seconds was converted to a speed measure by a reciprocal transformation and the reciprocals were averaged over successive trials (sets of five responses). The time measures for wrong responses were discarded, but there were very few of these after the first trial. Performance curves drawn from these data are shown in Fig. 1 and the results of an analysis of variance are set out in Table 1. The three main effects of practice, activity, and delay were significant, along with the double interactions of practice with activity and delay. The mean score for the first response in each group was included in Fig. 1 as a check on the comparability of the groups at the outset of training.

TABLE 1

ANALYSIS OF VARIANCE OF MEAN RESPONSE SPEEDS

Source	*df*	*MS*	*F*
Between *S*s	79		
Degree of activity (B)	1	9409.00	21.92**
Delay interval (C)	1	3306.25	7.70*
B × C	1	104.04	
Error	76	429.15	
Within *S*s	320		
Trials (A)	4	7735.61	369.77**
A × B	4	148.05	7.08**
A × C	4	108.77	5.20**
A × B × C	4	8.71	
Error	304	20.92	
Total	399		

* $P < .01$.
** $P < .001$.

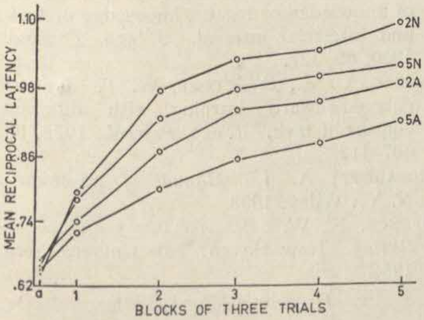


FIG. 1. Performance curves for groups learning with 2 and 5 sec. delay of reinforcement, and with activity (A) or inactivity (N) during the delay period. (The scores for Point a on the abscissa were taken from the the first item presented in Trial 1.)

DISCUSSION

Under the highly circumscribed conditions of this experiment with human *S*s results have been obtained which tally with those of Carlton's (1954) study using rats; since activity during delay of reinforcement impaired performance then Spence's hypothesis may be said to have been confirmed again. It might be argued that a strict application of the competing-response hypothesis should imply no impairment of performance with no activity during delay of reinforcement, so that Groups 2N and 5N should not have differed. The obtained superiority of Group 2N, however, may be ascribed to the technical impossibility of rendering *S* completely inactive, even for brief intervals (Bilodeau & Ryan, 1960), so that the designations A and N in this study refer to high and low degrees of activity rather than to its presence and absence, respectively. A further point of possible criticism concerns the interval between presentations. No attempt was made to hold this interval constant and it probably averaged 3 sec. longer for the 5-sec. delay groups. Bilodeau and Bilodeau (1958) have found some evidence for superior performance with shorter intertrial intervals as compared with a 7-day interval, but they found no significant differences with intervals of 20 sec., 2 min., and 1 hr. Denny and associates (Denny, Allard, Hall, & Rokeach, 1960) have reported significant differences with intervals of 10 sec. vs. 30 sec. and 20 sec. vs. 30 sec., but these *E*s applied no overall test of significance to their data before proceeding to compare groups in pairs. It thus seems reasonable to conclude that the possible difference between intervals in this study, of the order of 19 sec. vs. 22 sec., could have had no appreciable effect on performance.

In view of the positive results of the present study, despite the use of human *S*s, the conflict with the null results of delayed knowledge of results remains a problem, but the apparent discrepancy may depend on a false analogy between the two situations. In the experiment under consideration and in previous similar work with lower animals, the reinforcing event cannot be regarded as providing knowledge of results, for there has been no question of the correctness of the response, only its strength; a methodological concomitant of this fact has been the study of simple responses and the use of reaction latency or speed as a sensitive measure of response strength. In the other experiments with human *S*s, however, responses requiring some skill in their execution have been studied, the learner has been given some indication of the correctness of the response, and relatively crude or complex measures, such as frequency of correct response, have been used. The present *E*s sought to maintain the conditions of Spence's hypothesis with humans; if the process were to be reversed, so that rats served as *S*s in the usual type of study manipulating knowledge of results, then rather different conditions should obtain. For example, the rat as *S* might be required to learn to exert a certain degree of pressure on the lever, amount of reward could be scaled to the proximity of the obtained degree of pressure to that required, and frequency of correct pressure or "error" of pressure would then presumably be used as measures of performance. Under this latter set of conditions it could well prove that delay of reward has no measurable effect with lower animals, as has been the case with humans.

SUMMARY

With speed of response as a measure it was found that both delay of reinforcement and activity during delay had detrimental effects in a very simple learning situation with human *S*s. This result confirmed data from a similar study using rats and was interpreted as further supporting Spence's (1956) hypothesis about delay effects in terms of competing responses. Experiments on delay of knowledge of results in human motor learning, yielding no such effects, were seen to differ in certain basic respects from the present type of study.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

BILODEAU, E. A., & RYAN, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *J. exp. Psychol.*, 1960, **59**, 414–419.

CARLTON, P. L. Response strength as a function of delay of reward and physical confinement. Unpublished master's thesis, State University of Iowa, 1954.

CHAMPION, R. A. Reinforcement and learning theory. *Aust. J. Psychol.*, 1960, **12**, 10–20.

DENNY, M. R., ALLARD, M., HALL, E., & ROKEACH, M. Supplementary report: Delay of knowledge of results, knowledge of task, and intertrial interval. *J. exp. Psychol.*, 1960, **60**, 327.

NOBLE, C. E., & ALCOCK, W. T. Human delayed-reward learning with different lengths of task. *J. exp. Psychol.*, 1958, **56**, 407–412.

ROSANOFF, A. J. *Manual of psychiatry.* N. Y.: Wiley, 1938.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received May 1, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 593–600

THE DISTRIBUTION OF RECALLED ITEMS IN SIMULTANEOUS INTENTIONAL AND INCIDENTAL LEARNING[1]

ARNOLD MECHANIC[2]

University of California, Berkeley

Two types of incidental learning situations may be distinguished: Type I in which *S* is not instructed to learn any of the materials but is later tested; Type II in which *S* is given a circumscribed learning task and is later tested for materials to which he was exposed during learning but which are not covered by the learning instructions. In both situations, *S* is introduced to the incidental stimulus materials by means of an orienting task. This is a task which *S* is required to perform in order to insure that he is exposed to the incidental stimuli. The activities required by differing orienting tasks may vary widely as long as they insure *S*'s exposure to the incidental materials.

If instructions to learn generalize to materials not included in the instructions, amount of incidental learning should be greater in the Type II than in the Type I situation. However, such effects of generalization were not found in two previous studies using nonsense syllables. On the contrary, these studies, which will be reported in a forthcoming publication, indicated that the Type I situation produces significantly greater incidental learning. This was found to be true regardless of whether the intentional and incidental items were presented simultaneously or in temporal succession. If there were positive effects of generalization of instructions in these experiments, it would seem that they were completely countered by the negative effects of "task competition" in the Type II situation. By "task competition" we refer to the fact that the two kinds of material —the prescribed and the incidental— compete in terms of how much of the total exposure time is spent in responding to each kind of item. When *S*s are instructed to learn specific items, more of the available time will be utilized in responding to these items during exposure and in possible rehearsal thereafter. Consequently less time (or possibly no time) will be available for responding to the incidental items. The experiment to be reported in this paper studied task competition by manipulating the difficulty (meaningfulness) of both the intentional and incidental materials in a Type II situation. In addition, the effects of number of presentations were studied.

A review of the literature indicates no verbal learning data on Type II incidental learning as a function of intentional task variables. However, studies with nonverbal materials indicate that Type II learning varies with manipulations of the intentional task. Bahrick (1954) and Bahrick, Fitts, and Rankin (1952) found that the amount of learning irrelevant to a set is inversely related to the strength of the incentive determining that set. It might, therefore, be expected that varying the difficulty of the intentional items should also make for variations in the competition between the intentional and incidental tasks. Difficult intentional items could be expected to take up different proportions of *S*'s total learning time than would easy intentional items.

---

[1] Based upon a thesis submitted to the Graduate School of the University of California, Berkeley, in partial fulfillment of the requirements for the PhD degree. The author is greatly indebted to his major professor, L. J. Postman, under whose direction this research was carried out and to B. J. Underwood for his suggestions in the preparation of this manuscript. This research was conducted during the author's tenure on a research fellowship (MF-9934) from the National Institute of Mental Health of the United States Public Health Service.

The manuscript was prepared during the author's tenure as a National Science Foundation Postdoctoral Fellow at Northwestern University.

[2] Now at Alameda State College, Hayward, California.

Both meaningfulness and degree of practice have been studied in the Type I situation. The studies varying meaningfulness indicate that the superiority of intentional over incidental learning decreases as the meaningfulness of the items increases (Postman, Adams, & Phillips, 1955; Postman & Phillips, 1961). These findings were explained in terms of the differential responses which *S*s make to the stimulus items. More specifically, differential responses may be regarded as habits attached to a stimulus item either through previous experience with that item or through stimulus generalization. It is assumed that learning—both intentional and incidental—depends on the performance of these differential responses to the stimuli during exposure. In this view, the major effect of learning instructions is to increase the frequency and intensity of such differential responses. It may also be assumed that the high-meaningful items will evoke strong differential responses at the outset, while the low-meaningful items will not readily evoke such responses.

As a result of the strong pre-experimental habits attached to them, many of the high-M items will be learned in the absence of learning instructions. In other words, these items should be learned quite easily by both incidental and intentional learners. The low-M items, on the other hand, could be expected to evoke only weak responses in the absence of learning instructions. Instructions to learn would be required in order to strengthen these responses to the degree that they are capable of mediating learning. Consequently, the lower the meaningfulness of the materials, the greater should be the effects of instructions to learn.

TABLE 1

STIMULUS LISTS

List A—Low	List B—Low	List C—High	List D—High
DUR	OAC	WHI	ANC
LAK	TOS	SHO	DRE
DOP	BAK	ATI	LAN
HOK	LDE	TES	GRA
FUM	DEL	OUN	ITY
HEO	LOA	HAN	ILD
ERF	KDE	RON	BLE
RHA	SYC	WER	UST
COF	WIC	UND	OVE
WYE	KAB	NES	TER
SIL	EJO	OME	OUS
ARP	MPL	EST	RES

In the Type II situation, each *S* is exposed to both intentional and incidental lists. If the differential response hypothesis is correct, the difference between intentional and incidental scores for each *S* should decrease as the meaningfulness of the items increases.

The Type I data on number of presentations indicate that this variable has only marginal effects on the number of incidental items retained (Postman & Adams, 1958). Intentional learners showed a significant upward trend as a function of presentations whereas incidental learners did not. The present experiment will provide additional data on incidental learning as a function of presentations.

METHOD

Materials.—Meaningfulness was defined in terms of language frequency of trigrams (three-letter sequences not forming words). The learning materials consisted of four 12-unit trigram lists. Two of these lists were of very high-frequency items while the other two were of very low-frequency items. Two of the lists, one high-frequency and one low-frequency, were taken from Underwood and Postman (1960). The other two lists were constructed so as to approximate the characteristics of the Underwood and Postman lists. Items were selected on the basis of the Underwood "total" frequency count. The manner in which these frequencies were derived is described by Underwood and Schulz (1960). The four lists are shown in Table 1.

There is very little overlap in frequency between the low and high lists. On the basis of frequency per trigram in a sample of 1,035,000 English words, the frequency ranges are as follows: List A, 2–188; List B, 1–158; List C, 121–799; and List D, 127–1094. The approximate median frequencies are 5, 3, 350, and 375, respectively. None of the 48 trigrams appears in the Thorndike-Lorge (1944) word count.

The construction of these lists was governed by two criteria in addition to frequency. The first was formal intralist similarity defined in terms of duplication of letters. These lists were fairly well equated on this variable with the number of different letters for Lists A, B, C, and D being 17, 16, 13, and 16, respectively. The lower total number of letters used in List C was due to greater repetition of letters between first and second and between second and third positions.

The second criterion governing selection of items was the formal interlist similarity for all possible pairs of lists. Equating the pairs of lists on this variable was difficult because pairs of like frequency lists (H-H or L-L) tended to have more letter duplications than did pairs of mixed frequency lists (H-L or L-H). This is in all likelihood a function of different distributions of individual letters in the two populations (high-frequency trigrams vs. low-frequency trigrams). Therefore, the unmixed list pairings would be restricted to a narrower range of letters than would be the mixed pairings. Nevertheless it was possible to equate interlist similarity within reasonable limits. Duplicated letters between lists ranged from 9 to 13. Duplications in Position 1 ranged from 4 to 6; in Position 2, from 2 to 6; in Position 3, from 3 to 6. The unmixed list pairings averaged 5.5, 5.5, and 5 duplications in the three positions, respectively. The corresponding averages for the mixed list pairings were 5, 3.5, and 3.25.

Eight random orders were prepared for each of the four lists. Two of the lists were presented concurrently to each *S* during training. Each exposure of the lists involved presenting *S* with 12 pairs of items. Each pair was arranged vertically with one member in the top position and the other member directly below it. This arrangement was used to prevent the left to right chaining which might result from the reading habits of *Ss*.

The top members of the pairs made up one of the stimulus lists while the bottom members made up the second list. Successive presentations of the paired lists involved different random orders for each list. As a result, the pairings of individual items from the two lists were randomly varied from trial to trial. Stimulus materials were prepared for each of the six possible pairings of the four lists. Two variations for each pairing were prepared. These reversed the top and bottom member positions but were otherwise identical.

Conditions of learning.—All *S*s were instructed to learn one of the two lists presented and were discouraged by the instructions from attempting to learn the second list. Different groups were exposed to each pair of lists for two, four, or eight presentations. The item pairings were varied from trial to trial in order to prevent learning of the incidental items through the mediation of the intentional items. The experiment was introduced as a test of the notion that words meaning the same thing in different languages tend to sound more alike than words with different meanings. The *S*s were asked to judge the phonetic similarity between the members of each pair of trigrams. These were said to be items from two different primitive languages. The *S*s were required to pronounce the members of each pair to themselves and to rate their phonetic similarity on a five-point scale. They were instructed that the top members of the pairs were always from the same language while the bottom members were always from the other language. In order to get intentional learning of only one list, *S*s were told that we also wanted to study the ease of learning of one primitive language under conditions of distraction from a second language. In this way, instructions to learn were given for either the top or bottom items exclusively.

The *S*s were given a number of paper sheets, each covered with a strip of cardboard. A window was cut out of each cardboard strip in order to allow exposure of one pair of syllables at a time in accordance with *E*'s instructions. The pairs of syllables on each sheet were numbered 1 through 12. When *E* called out the number "1," *S* slid down the window to expose the first pair of syllables and entered his rating next to the syllables. The procedure was repeated for each successive pair until all 12 pairs on the page had been rated. The pairs were presented at a 10-sec. rate, with 20 sec. between presentations of the pages. A 5-min. test of free recall was given 1 min. after the end of practice. The *S*s were requested to write down as many of the syllables as they could remember regardless of whether they appeared in the top or bottom positions of the pairs. They were told to write the syllables down in the order they were remembered, and that their scores

TABLE 2

AVERAGE NUMBER OF INCIDENTAL ITEMS RECALLED BY THE VARIOUS GROUPS

Presentations	Incidental Learning Items			
	High Meaningful		Low Meaningful	
	High Int.	Low Int.	High Int.	Low Int.
2	2.31	2.25	1.06	1.44
4	3.56	4.06	2.31	2.62
8	4.94	4.69	3.88	3.19

would depend only on the number of correct items that they wrote down. When the recall test was completed, a postexperimental inquiry was made to ascertain whether *S* understood the instructions and whether, for any reason, an attempt was made to learn the incidental items. Each *S* received two learning scores, one being the number of correctly reproduced intentional items and the other being the number of correctly reproduced incidental items.

Experimental design.—Twelve groups of 16 *S*s each were run in a standard 2 × 2 × 3 factorial design. The three independent variables were meaningfulness of the intentional items, meaningfulness of the incidental items, and number of presentations. The design was balanced for top and bottom order of lists, and for the position of the intentional items. With an overall total of six possible list combinations, two possible sequences of the items from the two lists within the pairs (i.e., A-B or B-A), two possible positions for the intentional items, and three degrees of practice, a total of 72 task variations were required for the balanced design.

Subjects.—The *S*s were students in lower division psychology courses at the University of California. They did not know the purpose of the experiment and were assigned to the conditions by rotation. They served in groups, averaging 13 in number. Each of these groups included several different experimental conditions for which the instructions were identical but which differed with regard to the stimulus lists given to *S*s. Of necessity, two-presentation, four-presentation, and eight-presentation conditions had to be run separately. Otherwise, the experimental conditions were assigned by rotation to the members of each laboratory group with correction for equal *N*s. The two-presentation, four-presentation, and eight-presentation treatments were assigned to the groups by rotation.

A total of 237 *S*s were tested in order to obtain the 192 required by the experimental design. Of these, 22 indicated upon inquiry that they did not understand the instructions or that they made some attempt to learn the incidental items. They were discarded and replaced. Of these, 7 were in the two-presentation group, 8 in the four-presentation group, and 7 in the eight-presentation group. In filling out the laboratory groups during the final phase of the experiment, 23 extra *S*s were run. Where there were extra *S*s for any of the 72 task variations, the *S*s to be used in the analysis were selected randomly.

RESULTS AND DISCUSSION

Incidental learning.—The incidental learning scores were analyzed in terms of the 2 × 2 × 3 factorial design. The mean numbers of correctly reproduced incidental items are given in Table 2. For example, the first entry indicates that an average of 2.31 high-M items were learned incidentally, where the concurrent intentional items were also high-M, and there were two presentations. Table 3 summarizes the analysis of variance for the incidental learning data presented in Table 2. Because Bartlett's test showed significant heterogeneity of variance, the Freeman-Tukey square-root transformation ($\sqrt{x} + \sqrt{x+1}$) was used in this

TABLE 3

ANALYSIS OF VARIANCE OF INCIDENTAL SCORES

Source	*df*	*MS*	*F*
Meaningfulness (incident. items) (M_1)	1	26.85	21.46*
Meaningfulness (intent. items) (M_2)	1	.00	.00
No. of presentations (P)	2	35.25	28.17*
$M_1 \times M_2$	1	.04	.03
$M_1 \times P$	2	.29	.23
$M_2 \times P$	2	.53	.42
$M_1 \times M_2 \times P$	2	.15	.12
Error	180	(1.25)	—

* $P < .001$.

analysis, as well as in the analysis of intentional scores below. In both cases, Bartlett's test indicated that there was no significant heterogeneity in the transformed scores. It can be seen that meaningfulness of the incidental items, and number of presentations are highly significant sources of variance. The meaningfulness of the concurrently learned intentional items is not a significant source of variance in incidental learning. None of the interactions among the independent variables are significant.

If task competition is an important variable in Type II incidental learning, it might be expected that incidental learning would vary as a function of intentional-task difficulty. Therefore the question may be raised as to why incidental learning did not vary with the meaningfulness of the intentional items. The failure to find a relationship between intentional-M and incidental learning is surprising in view of the nonverbal studies which indicate that Type II incidental learning varies significantly as a function of intentional task variables (Bahrick, 1954; Bahrick et al., 1952). The considerably greater effects obtained in the nonverbal experiments become explicable when we consider the differences between the verbal and nonverbal situations. These situations differed greatly with regard to the activities required by the orienting tasks. These activities may facilitate or inhibit the performance of those differential responses which are assumed to be necessary for learning (Postman & Adams, 1956). Pronunciation of verbal items probably makes for very adequate differential responses as far as learning is concerned. In the verbal situation, the orienting task required that *S* make definite pronouncing responses to both the intentional and incidental items. Because of the nature of the orienting task, task competition could not prevent *S* from producing differential responses to the incidental items. At most, task competition could only reduce the frequency and intensity of *S*'s responses. In the case of the nonverbal experiments cited above, the situation was quite different. There was no specific orienting task which required *S* to respond differentially to the incidental stimuli. It was possible for *S* to completely ignore these stimuli while still conforming to the instructions of the experimenter. For example, *S* should be able to learn geometric forms intentionally without necessarily responding to their incidental colors. Therefore, large amounts of competition from the intentional task could result in failure to perform differential responses to the incidental materials.

It is evident from the foregoing analysis that the orienting task can be an important factor in determining the effects of task competition on incidental learning. Where the orienting task merely requires stimulus exposure of the incidental materials, the role of task competition can be great. But where the orienting task requires differential responding to the incidental materials, the role of task competition may be severely limited. In view of these considerations, it is not surprising that Bahrick's manipulations of the intentional task produced greater effects on incidental learning than did ours. We should predict that the effects of incentive upon Type II incidental learning would be sharply reduced if the orienting task required differential responses to the incidental items. Similarly, the effects of intentional-M on incidental learning should be increased by use of a less facilitating orienting task. For example, *S* might be required to assign "guessed" numbers to each incidental item instead of pronouncing it. The degree of task competition might then become an important factor in determining whether or not *S* makes appropriate differential responses to the incidental items.

Intentional learning.—The design of the analysis of intentional scores was identical to the design for the incidental scores. The means for

TABLE 4

AVERAGE NUMBER OF INTENTIONAL ITEMS RECALLED BY THE VARIOUS GROUPS

Presentations	Intentional Learning Items			
	High Meaningful		Low Meaningful	
	High Inc.	Low Inc.	High Inc.	Low Inc.
2	5.69	5.19	4.12	4.50
4	7.94	7.25	7.25	6.81
8	8.81	9.19	9.00	8.62

each condition are given in Table 4. The only significant sources of variance are meaningfulness of the intentional items ($F = 4.72$, $df = 1/180$, $P < .05$) and number of presentations ($F = 49.07$, $df = 2/180$, $P < .001$).

Differences between intentional and incidental learning.—Postman et al. (1955) and Postman and Phillips (1961) have found that the superiority of intentional over incidental learning decreases as the meaningfulness of the items increases. In the Type II situation, each *S* is exposed to both intentional and incidental lists. It is of interest to ascertain whether the difference between intentional and incidental scores for each *S* will also decrease as the meaningfulness of the items increases. Such a finding would be consistent with the Type I studies which used different groups of *S*s for intentional and incidental learning. In order to test this hypothesis, differences between intentional items recalled and incidental items recalled were obtained for each *S*. These were then divided by the *S*'s total learning in order to obtain relative difference scores for each *S* $\left(\frac{\text{Int.} - \text{Inc.}}{\text{Int.} + \text{Inc.}}\right)$. This resulted in measures of the difference between intentional and incidental learning which were comparable for different levels of overall learning. Each *S* in the experiment received one of four possible list combinations. These combinations can be ordered in terms of predicted average differences between intentional and incidental learning. Such an ordering involves two assumptions. First, that increased meaningfulness increases learning. This assumption has already been borne out by the significant *F* ratios for meaningfulness in the separate analyses of incidental and intentional learning. The second assumption is that the difference between intentional and incidental learning will decrease as the stimulus items increase in meaningfulness. On the basis of these two assumptions, the intentional-incidental differences for the four list combinations should be in the following increasing order: (1) Low Intentional-High Incidental; (2) High Intentional-High Incidental; (3) Low Intentional-Low Incidental; and (4)

TABLE 5

AVERAGE RELATIVE DIFFERENCES BETWEEN INTENTIONAL AND INCIDENTAL LEARNING FOR THE VARIOUS LIST COMBINATIONS

	List Combinations			
	(1) Low Int. High Inc.	(2) High Int. High Inc.	(3) Low Int. Low Inc.	(4) High Int. Low Inc.
Two Trials	19.23	43.94	60.04	69.58
Four Trials	34.92	42.10	45.73	51.46
Eight Trials	31.41	28.35	47.66	41.55
All Trials (two, four, & eight combined)	28.52	38.13	51.14	54.20

Note.—Means are based upon 16 *S*s for each list combination at each level of practice. The score for each *S* is: $\frac{\text{Int.} - \text{Inc.}}{\text{Int.} + \text{Inc.}} \times 100$.

High Intentional-Low Incidental. A significant gap between the second and third list combinations would indicate that the difference between the two kinds of learning varies as a function of the meaningfulness of the items.

Table 5 gives the mean relative differences for each list combination at each level of practice. The bottom row of the table gives the average differences when the three levels of practice are combined. It can be seen from this row that the differences increase in the predicted order. The largest difference is between List Combinations 2 and 3.

The data for two and four trials are also in the predicted order without any reversals. At eight trials, there are small reversals between Combinations 1 and 2 and between Combinations 3 and 4. However, Combination 3 shows a larger difference than Combination 2 at all levels of practice.

The difference scores were subjected to an analysis of variance with list combinations and number of presentations as the independent variables. Because the scores in this analysis are percentages, the arc-sine transformation (Angle = Arcsine $\sqrt{\text{Percentage}}$) was used. As hypothesized, list combinations is a highly significant source of variance ($F = 6.58$, $df = 3/180$, $P < .001$). The number of presentations does not contribute significant variance to the percentage-difference scores ($F = 1.78$, $df = 2/180$). The interaction between list combinations and presentations is also not significant ($F = 1.64$, $df = 6/180$).

In order to analyze further the significant effects of list combinations, the gap test of Tukey (1949) was utilized. The overall means of the transformed scores were tested in order to determine whether any of the differences between adjacent means were significant. These means differ from those in the bottom row of Table 5 only in that the scores have been transformed. The means for List Combinations 1–4 are 27.86, 35.16, 46.63, and 49.34, respectively. A gap as large as 10.85 between adjacent means is significant beyond the .05 level for a two-tailed test and beyond the .025 level for a one-tailed test. In view of the predicted ordering of conditions, the one-tailed probabilities are appropriate. It may be seen that the only gap larger than 10.85 ($P < .025$) is between List Combination 2 (H-H) and List Combination 3 (L-L). This indicates that the higher the meaningfulness of the items, the smaller is the difference between intentional and incidental learning.

The gap between List Combinations 1 and 2 falls short of statistical significance ($.05 < P < .10$). Therefore the application of Tukey's test divides the means into two significantly separated pairs: (*a*) the pair of mean differences when incidental items are of high meaningfulness; (*b*) the pair of mean differences when incidental items are of low meaningfulness. Differences between intentional and incidental learning are significantly greater in the case of the latter pair.

With regard to the differential response hypothesis, the crucial comparison of mean intentional-incidental differences is between the two groups with unmixed lists (H-H vs. L-L). As indicated above, the intentional-incidental difference is significantly greater when both stimulus lists are of low meaningfulness than when both are of high meaningfulness. This is a confirmation of the hypothesis using intentional-incidental differences for individual *S*s. Previous supporting data from Type I experiments have involved comparisons of groups of intentional learners with groups of incidental learners.

The role of practice.—Another finding of this experiment was that both intentional and incidental learning increased reliably as a function of presentations. These data are in contrast to a previous study on Type I incidental learning (Postman & Adams, 1958). There it was found that intentional learning showed a significant upward trend as a function of presentations whereas incidental learning did not. These different findings are probably due to the difference between orienting tasks in the two experiments. It has been shown that differences between intentional and incidental learning vary as a function of the nature of the orienting task (Postman & Adams, 1956). The earlier study of the effects of presentations required that *S* guess numbers assigned by chance to each syllable. *S*s were told that each presentation of the series would have different number assignments. The *S*s were faced with relatively novel nonsense syllables during the initial presentations of the series. With repeated presentations, there was no requirement for incidental *S*s to respond to the syllables differentially. Therefore it might be expected that practice would have only marginal effects on incidental learning while having large effects on intentional learning. The present experiment required that *S* pronounce the items to himself in order to rate phonetic similarity. Because this form of response to the incidental items was required of *S* on every presentation, it is not surprising that incidental learning increased with practice in very much the same manner as did intentional learning.

SUMMARY

This experiment studied "task competition" by manipulating the difficulty of both the intentional and incidental materials in a Type II incidental learning situation, in which *S* is exposed to two sets of materials, instructed to learn only one of the sets, and is later tested for the materials which he was not instructed to learn.

Twelve groups of 16 *S*s each were run in a standard 2 × 2 × 3 factorial design. The three independent variables were meaningfulness of the intentional items (high vs. low), meaningfulness of the incidental items (high vs. low), and number of presentations (two, four, eight). Scores for both incidental and intentional learning were obtained from each *S*.

The first analysis of the data studied incidental learning scores in terms of the 2 × 2 × 3 factorial design described above. The meaningfulness of the incidental items was a significant source of variance in incidental learning. The meaningfulness of the intentional items did not contribute significant variance either by itself or in interaction with the other independent variables. In contrast with an earlier study, it was found that incidental learning increased significantly as a function of practice.

The second analysis of the data involved the difference scores between intentional and incidental learning for each *S*. It was predicted that differences would be greater when the items were of low meaningfulness than when they were of high meaningfulness. Analogous results have been found in the Type I situation and have been explained in terms of a differential response hypothesis. The predicted results in this Type II experiment support such an hypothesis.

REFERENCES

BAHRICK, H. P. Incidental learning under two incentive conditions. *J. exp. Psychol.*, 1954, **47**, 170–172.

BAHRICK, H. P., FITTS, P. M., & RANKIN, R. E. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, **44**, 400–406.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: IV. The interaction of orienting tasks and stimulus materials. *J. exp. Psychol.*, 1956, **51**, 329–333.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: VII. Effects of frequency of exercise and length of list. *J. exp. Psychol.*, 1958, **56**, 86–94.

POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and of the method of testing. *J. exp. Psychol.*, 1955, **49**, 1–10.

POSTMAN, L., & PHILLIPS, L. W. Studies in incidental learning: IX. A comparison of the methods of successive and single recalls. *J. exp. Psychol.*, 1961, **61**, 236–241.

THORNDIKE, E. L., & LORGE, I. *A teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received May 2, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 601–609

CONSISTENCY AND COMPLEXITY OF RESPONSE SEQUENCES AS A FUNCTION OF SCHEDULES OF NONCONTINGENT REWARD[1]

JOHN C. WRIGHT[2]

Stanford University

It is frequently noticed in experiments on learning that certain nonessential aspects of a response, or even wholly nonessential responses, are reinforced along with those responses or aspects of response upon which reinforcement is contingent (Guthrie & Horton, 1946; Morse & Skinner, 1957; Skinner, 1948).

Also, when analysis is made of orderly response tendencies as a function of random, noncontingent reinforcement of such tendencies, it is found that *S*s respond in an orderly fashion, though the patterns generating such orderliness vary greatly both within and between *S*s.[3] For example, in an arithmetic concept formation task, Jenkins (1959) found that the higher the probability of noncontingent reinforcement, the more individual *S*s tended to converge upon a single operation or rule for generating their responses. He noted that rules, rather than responses, seemed to be established by this process. Since the operations "learned" were in fact superstitious and bore no contingent relationship to reinforcement, it was impossible to predict what operation a particular *S* would converge upon, but the degree of convergence appeared to be regularly controlled by the probability of noncontingent reinforcement selected by *E*.

The present experiment is an investigation of the effect of the probability of noncontingent reward on the development of response sequences in human *S*s.[4] The purpose is to determine how much of the process of developing and testing hypotheses is determined simply by the occurrence of reward, regardless of what response aspects happen to precede reward.

METHOD

Apparatus

The *S*s were seated before a plain, black, sloping panel. Centered on the panel at chest level was a circle of 16 regularly spaced, black pushbuttons. A seventeenth button was mounted in the center of the circle. All buttons were unmarked and identical except for location on the panel. The buttons were 1.8 cm. in diameter, and the diameter of the circular array was approximately 20 cm. Above the sloping panel was a vertical panel containing a three-digit, cumulative counter visible to *S*.

There were two independent circuits within the apparatus. The recording circuit consisted of an electric typewriter, concealed from *S*, with 16 of its keys arranged so that each could be activated by a different button. Thus a code letter was recorded on paper each time *S* pressed one of the 16 buttons on the perimeter of the circle. The typewriter was wired so that the platen rolled up one line and the carriage returned automatically following each block of 25 recorded responses. The *S*s operated the buttons continuously, generating a typed record of the response sequence, with 25 characters in each line. On responses followed by reward, the space bar on the typewriter was activated, thereby yielding a record of the specific responses that happened to be rewarded, as well as a confirmation that the preprogramed sequence of rewarded and unrewarded trials was delivered as scheduled.

---

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree in the Department of Psychology of Stanford University. The author wishes to acknowledge the advice and support of Alex Bavelas in the design and conduct of this experiment.

[2] Now at the University of Minnesota.

[3] A. Bavelas, personal communication, 1958.

[4] The term, *reward*, is used in place of reinforcement in order to distinguish the present situation from that in the Humphreys-Estes guessing situation (Estes & Straughan, 1954). In the guessing situation, *S* is instructed to *anticipate* which of two events will occur, while in the present situation *S* is to *control* which of two events occurs, and to maximize the occurrence of one of them. In the guessing situation every outcome is said to be a reinforcing event; in the present situation rewards are reinforcing, and nonrewards are not.

The event-scheduling circuit consisted of a 250-position stepping switch, wired to a control panel so that any of the 250 consecutive positions on the stepping switch could be activated by a single wire on the panel. Thus, a consecutive string of 250 trials could be preprogramed as to their outcomes. On each trial *S* was instructed to press one of the 16 perimeter buttons and then to press the center (seventeenth) button to "test" whether or not his preceding response had earned him a point. In fact, the center test button merely advanced the stepping switch to the next position, where, if the corresponding hole in the control panel had been wired, *S* received a reward, consisting of an additional point on the counter and a buzzing sound for the duration of the test response. On trials when the panel had not been wired for a reward, the depression of the test button produced only the click of the advancing stepping switch and no additional points on the counter.

TABLE 1

SCHEDULES OF NONCONTINGENT REWARD FOR THE FIVE GROUPS

Group	Block Number	Value of π	Number of Rewarded Trials per Block
20	1 to 10	.20	5
50	1 to 10	.50	12 on odd numbered blocks 13 on even numbered blocks
80	1 to 10	.80	20
Rising	1	.04	1
	2	.08	2
	3	.20	5
	4	.32	8
	5	.44	11
	6	.56	14
	7	.68	17
	8	.80	20
	9	.92	23
	10	.96	24
Falling	Same as rising, but in reverse order		

Procedure

The *S*s were recruited from the introductory psychology course at Stanford University, to take part in "an experiment in problem solving." Eighty *S*s were randomly assigned to five groups. The sex distribution was ignored. The *S*s were instructed to operate the buttons, one at a time, so as to maximize the number of buzzes, indicated by the counter. They were told that one or more correct selections, patterns, combinations, or sequences of buttons would be rewarded. Each response was to be tested by pressing the center button before the next response was made. The *S*s were told to work at a steady, comfortable pace, since the duration of the experiment was determined by the number of responses rather than elapsed time.

All *S*s then operated continuously and without comment or interruption by *E* for 325 trials, taking 30–40 min. to do so. The trials were divided into 13 blocks of 25 trials each. During Blocks 11 and 12, no rewards were given, and in Block 13 every response was rewarded. These two conditions in the last three blocks were the same for all groups, and will henceforth be referred to as "extinction" and "recovery," respectively. It was felt that 0% and 100% reward were special conditions that might have discontinuous effects on response sequences, and accordingly they were applied only after all the other conditions in order to obtain rough estimates of behavior at these extreme values.

Blocks 1 to 10, the first 250 trials, contained the experimental variations. In the following description π is defined as the probability of reward, and $\pi = .20$, say, means that within the block in question exactly 20% of the trials were rewarded, regardless of *S*'s responses, the sequence being randomized once for all *S*s in that group on that block. Whenever the same π value was called for in a different group of *S*s, or in a different block for the same group, the sequence of rewarded and unrewarded trials was again randomized. Five groups of 16 *S*s each were run under different schedules of noncontingent reward prior to extinction and recovery, as shown in Table 1. The first three groups will be referred to as the "fixed groups" because for them π was constant throughout the first 10 blocks. The other two groups will be referred to as the "sliding groups" because the value of π changed in each successive block.

The purpose of the sliding group schedules was to determine whether regular changes in π would produce any special effects beyond those due to the various π values themselves. Only nonsignificant suggestions of such effects were obtained, and therefore only effects due to π values, regardless of order of presentation, will be considered.

Scoring the Results

The nature of the experiment required that an entirely new method of scoring be developed, one that would be equally sensitive to orderliness in responding, regardless of what peculiar and complex hypothesis an individual *S* used to generate a response sequence. One cannot assume that the scorer's ingenuity is always equal to or greater than that of *S*. Nor can one get uniformly intelligible verbal descriptions of rules for generating response sequences from *S*s themselves. Consequently, a scoring system was devised to measure any response sequence on a dimension from orderliness and predictability to randomness and unpredictability. The measure finally obtained for various transformations of the scores is called *response uncertainty* and is measured in runs per block. With the restrictions noted below, it is a rough measure of randomness in response sequences. It is inversely analogous to the "transmitted information" measure of Shannon (1948) and to the "stereotypy" measure of Miller and Frick (1949).

Zero-order scoring.—The raw data were in the form of continuous response chains, the responses being selected by *S* from among the 16 response buttons. The buttons were arbitrarily numbered clockwise from 1 to 16, and each response was recorded as the number of the button pressed. The *S*s had no knowledge of this arbitrary notation. In each block of 25 consecutive responses, the zero-order score was defined as the number of runs of a single button in each block. A minimum score of 1 run would indicate that *S* pressed a single button 25 times in succession in that block, and was maximally consistent. A maximum score of 25 runs in a block of 25 responses would indicate that *S* never pressed the same button twice in succession and therefore was minimally consistent in that block. The number of runs at this level of scoring is called zero-order response uncertainty. It measures inconsistency of responding, with the restriction that consistency, at the zero-order level of scoring, can result only from a tendency to repeat responses, or to eliminate responses from consideration. It is assumed that at this level restriction of the range of responses and the tendency toward runs on a single response are part of the same process of convergence.

First-order scoring.—Most *S*s converged on response rules rather than on specific responses. The most common example is constant-interval rotation. In order to detect the consistent use of such rules, the response chain of each *S* was transformed as follows: Each response was rescored by counting clockwise the number of steps (buttons in the array) required to move from that response to the next. The first-order transformation for each response indicates how far around the circle of buttons *S* "moved" from each response to the next in succession. Each block of 25 responses yielded 24 such intervals. The twenty-fifth interval for each block was defined as the interval between the last response on that block and the first response on the next block. Using the transformed data, first-order response uncertainty was defined as the number of runs of a single interbutton interval in one block. A score of 1 would indicate that the distance between successive responses was constant, and therefore that *S* was maximally consistent. Conversely, a score of 25 would indicate that *S* never used the same interbutton interval twice in succession, and was therefore minimally consistent. A pattern of consistent button repetition (low zero-order response uncertainty) would also appear as consistent in first-order scoring, since for button repetition, every interbutton interval is zero.

Second-order scoring.—A few *S*s converged upon response rules other than repetition or constant-interval rotation, so that consistency would not be detected by either zero- or first-order scoring. Therefore, the intervals produced by first-order scoring were again transformed. The 16 intervals were reduced to 9 by taking the shortest interval between responses, regardless of rotational direction. The absolute difference between each interbutton interval and the immediately following interbutton interval yields a transformed sequence of 25 scores per block. The

TABLE 2

Illustration of Scoring Transformations and Measures of Response Uncertainty for a Hypothetical *S*

Order of Scoring	Scoring Transformations / ←———————— one block of 25 trials ————————→ / ←—— next block	Response Uncertainty[a]
Zero	/2 2 2 2 2 2 2 3 4 5 6 7 8 9 11 13 15 1 3 4 6 9 13 2 8 / 15 5 10	19 runs
First	/ 0 0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2 2 1 2 3 4 5 6 7 / 6 5	10 runs
Second	/ 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 1 1 1 1/ 1	6 runs

Note.—Rewards are not indicated in the table.
[a] In runs. Each run is underlined.

number of runs of these scores per block defines second-order response uncertainty. Again the range is from 1 (maximum consistency) to 25 (maximum inconsistency). Second-order scoring, like first-order scoring, is sensitive to response repetition and to response rules of constant interval rotation, but it can also detect consistent use of progressively changing interval rotation, single and double button alternation, and single interval alternation as response rules. Table 2 illustrates the scoring method and the response uncertainty values obtained at each level of scoring for a hypothetical *S* on a single block of 25 responses.

It can be shown that still higher orders of scoring would detect still more subtle and complex kinds of consistency in responding, but inspection of second-order transformations for all *S*s revealed that only 2 of the 80 *S*s made use of response patterns which, though apparent in their second-order scores, would require third-order scoring in order to be reduced to runs.

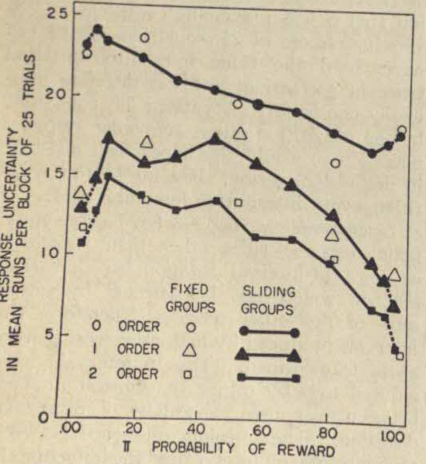


FIG. 1. Response uncertainty as a function of π.

Null values for each order of response uncertainty have been calculated for independent, random sequences of responses. Given the response on Trial $n - 1$, the probability of a response repetition on Trial n is simply 1/16. The expected number of repetitions per block is thus 25/16, or 1.56. The expected value of zero-order response uncertainty for a single *S* in a single block is thus 23.44. The same expected value holds for first-order response uncertainty, since given any of 16 possible intervals between Response $n - 1$ and Response n, then only 1 of the 16 responses possible on Trial $n + 1$ will generate a repetition of that interval, thus reducing the number of runs in that block by one.

The null value for second- and higher-order response uncertainty is somewhat lower, because direction of rotation is ignored, yielding 9 rather than 16 possible intervals. Two of the 9 intervals (0 and 8) can be generated by only one response: 0 by a button repetition and 8 by selection of the diametrically opposed button. Each of the remaining 7 intervals can be generated by either of two responses, depending on the direction of rotation. The probability of a repeated interval change is thus $(2/16 \times 1/16) + (14/16 \times 2/16)$, or .117. The expected number of repeated scores per block is thus 2.93, and the expected value of second-order response uncertainty is 22.07, for a single *S* on a single block of trials.

Complexity scoring.—By subtracting second-order response uncertainty from first-order response uncertainty, it is possible to measure to degree to which *S*s used consistent response rules more complex than either repetition or constant interval rotation. The resulting score is called *complexity*, defined as that component of general consistency due exclusively to the use of the more complex patterns detectable only at the second-order scoring level. Under conditions of random response selection, the expected value for complexity is 1.37, the difference between the null values for first- and second-order response uncertainty.

RESULTS

Figure 1 shows response uncertainty (mean number of runs per block) as a function of π (probability of noncontingent reward) for zero-, first-, and second-order scores. Response uncertainty for the π values of 0 and 1 have been estimated from the extinction and recovery blocks, respectively, and are omitted from the statistical analyses.[5] Since the order of presentation of π values for the rising and falling groups is counterbalanced, and since response uncertainty for the two groups did not differ significantly, either overall or in interaction with π values, their scores have been combined in a single curve labeled "sliding groups."

Zero-Order Scores

There is an approximately linear, negative relationship between zero-order response uncertainty and probability of reward. Button repetition is greatest when π is highest. Response uncertainty for the fixed groups is matched by that for the sliding groups at corresponding π values. To test the effect of π on response uncertainty, analysis of variance was supplemented with trend analysis using orthogonal polynomials (Lewis, 1960). For the three fixed groups, the linear component of the between-groups variance due to π values is significant at the .005 level, and the residual is nonsignificant. For the rising and falling groups combined, only the linear component of variance due to π values is significant ($P < .001$). No test of the interaction between *S*s and π values is possible, and the interaction between π value and condition (rising vs. falling) is not significant. It is concluded that the tendency to repeat a button is a linear positive function of the probability of noncontingent reward.

First- and Second-Order Scores

Since successively higher orders of scoring take into account convergence (reduction in response uncertainty) due to the consistent use of progression rules, as well as that due to button repetition, it is evident that higher orders of scoring cannot yield higher response uncertainty, and will usually yield lower response uncertainty at all π values. This result is in part a product of the scoring scheme and is clearly indicated in Fig. 1, where higher order scoring yields lower response uncertainty.

The most striking finding is that the shift from zero-order scoring (for response repetition) to second-order scoring (for all kinds of response orderliness detected) produces a marked change in the shape of the π function. The function becomes nonmonotonic, having a flat maximum between .08 and .50. Below .50 response uncertainty does not rise, and as π approaches zero, uncertainty shows some indication of declining. Since the curve for second-order scores roughly parallels that for first-order scores and is a more general measure of convergence, only the analysis of second-order scores will

---

[5] Combined means for the fixed groups (20, 50, and 80) and combined means for the sliding groups (rising and falling) are used as estimates of two sets of extinction and recovery points. No significant differences among groups were found for the extinction or the recovery values of response uncertainty.

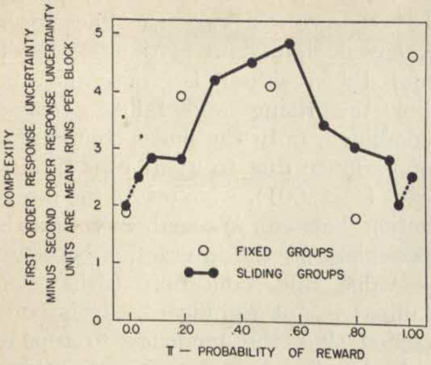


FIG. 2. Complexity of response rules as a function of π.

be reported. For the three fixed groups, the linear component of the between-groups variance is significant at the .025 level, and the residual is not significant. Therefore, on the basis of the three fixed groups alone, we may not conclude that the relationship is curvilinear. Doubtless significant curvilinearity would appear if the extinction and recovery estimates could legitimately be included in the analysis, to yield five data points instead of three. For the sliding groups, which again correspond rather closely to the fixed groups at the appropriate π values, the linear component of variance due to π values is significant at the .0001 level, and the quadratic component is significant at the .001 level. No other components approach significance. This means that there is a strong negative relationship between π and response uncertainty in general, and that there is also a highly significant curvilinear trend.

Complexity Scores

Figure 2 shows complexity as a function of π. As in Fig. 1, the rising and falling groups are combined. Again the extinction and recovery values are estimates and are excluded

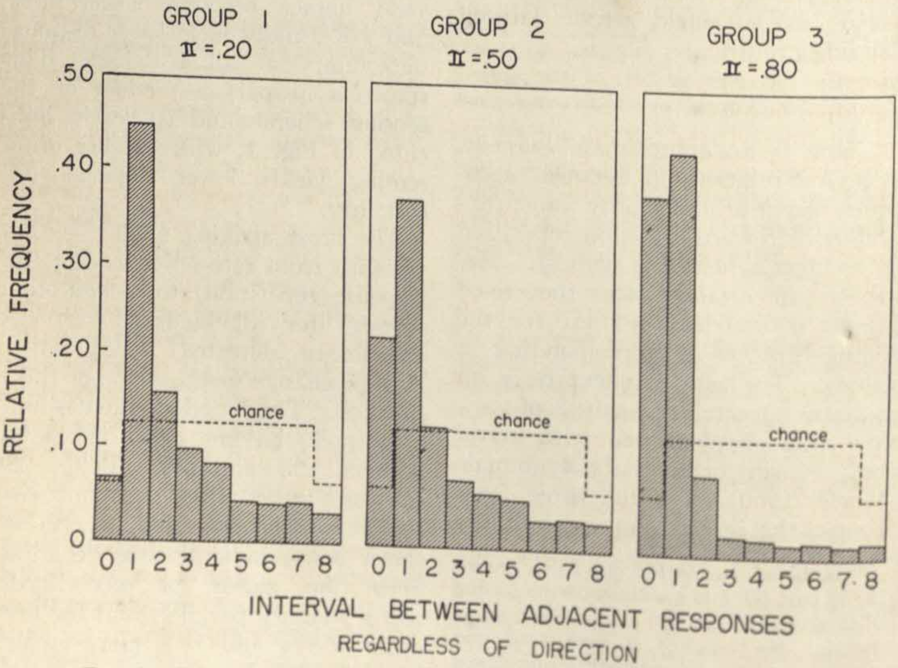


FIG. 3. Frequency distributions of response intervals, fixed groups only; $N = 16$ *S*s per group, 250 responses per *S*.

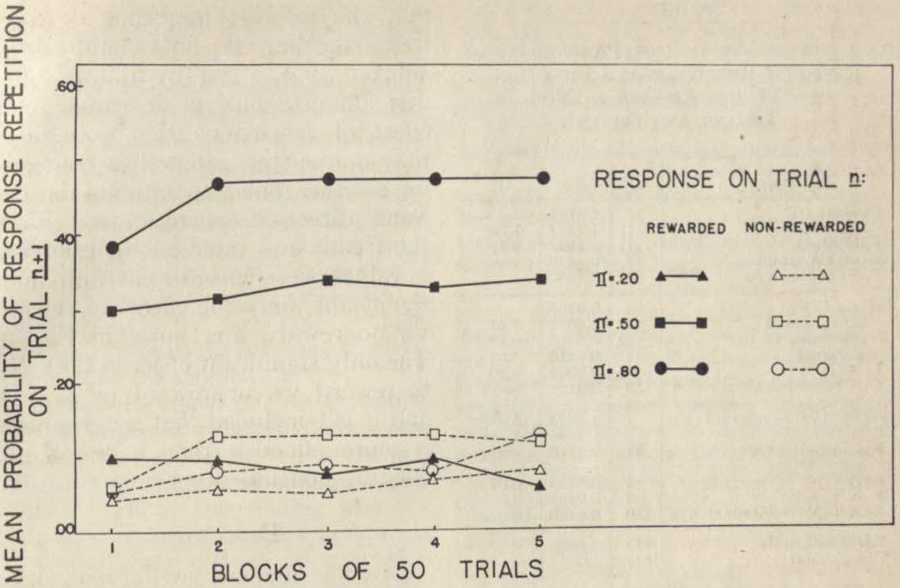


FIG. 4. Mean probability of a response repetition on Trial $n + 1$ as a function of reward and nonreward on Trial n, fixed groups only; in blocks of 50 trials.

from the analysis. Trend analysis of the complexity scores yields no significant trends due to π values for the fixed groups. For the sliding groups, however, π values contribute a quadratic trend significant at the .001 level, no significant linear trend, and no higher order trends, Therefore, there is a significant curvilinear relationship between π and complexity.

Distribution of Response Intervals

Figure 3 shows the relative frequency distributions of the intervals between successive responses for the three fixed groups, based on 4,000 trials per group. Clearly, intervals of zero (response repetition) and one (single step rotation) predominate. The only significant between-groups difference is in the relative frequency of response repetition, where the linear trend due to π values is significant at the .005 level.

Single Trial Effects

Repetition.—As π is increased, response repetition is increased. Analysis by single trials is summarized in Fig. 4, which shows the proportions of rewarded and nonrewarded trials that were followed on the next trial by response repetition, for the three constant groups over blocks of 50 trials. Not only does reward have a clearly reinforcing effect on the preceding response, but the magnitude of the effect varies positively with π. Table 3 summarizes the trend analyses for probability of response repetition. Note that both trends over trial blocks and trends over π values were tested. The reduction in zero-order response uncertainty as π increases is thus due not only to the increased number of rewards, but also to the increased effectiveness of each reward as a reinforcement for button repetition.

Rotation.—Although Fig. 3 shows

TABLE 3

ANALYSIS OF VARIANCE OF PROBABILITY OF RESPONSE REPETITION AS A FUNCTION OF π VALUES, REWARD VS. NONREWARD, AND TRIALS

Source	*df*	*MS*	*F*
π values (G)	2	1.6788	
Linear trend	(1)	3.1383	7.26**
Residual	(1)	0.2193	0.51
*S*s within groups	45	0.4320	error
Total between *S*s	47	0.4850	
Trials (T)	4	0.0265	
Linear trend	(1)	0.0675	3.67
Quadratic trend	(1)	0.0164	0.89
Residual	(2)	0.0110	0.60
T × G	8	0.0092	0.50
T × *S*s within groups	180	0.0184	error
Reward vs. Nonreward (R)	1	4.8823	18.93***
R × G	2	1.2240	4.75*
R × *S*s within groups	45	0.2579	error
R × T	4	0.0013	0.09
R × T × G	8	0.0105	0.75
R × T × *S*s within groups	180	0.0140	error
Total within *S*s	432	0.0580	
Grand total	479		

* $P < .05$.
** $P < .01$.
*** $P < .001$.

that the relative frequency of one-step rotation is not significantly affected by π, a trial-by-trial analysis was undertaken to determine the effect of reward on the probability that a one-step interval is followed by another one-step interval in the same direction of rotation. Since the trends and interactions involving π values were inconsistent and nonsignificant, only the effects of reward vs. nonreward are shown in Fig. 5. The only significant effect is that due to reward vs. nonreward ($P < .01$), and it is concluded that a nonreward is more effective than a reward in eliciting continued one-step rotation.

DISCUSSION

Given a minimum level of reward, *S*s will converge on and repeat consistently some response or rule as a function of the relative frequency of reward pro-

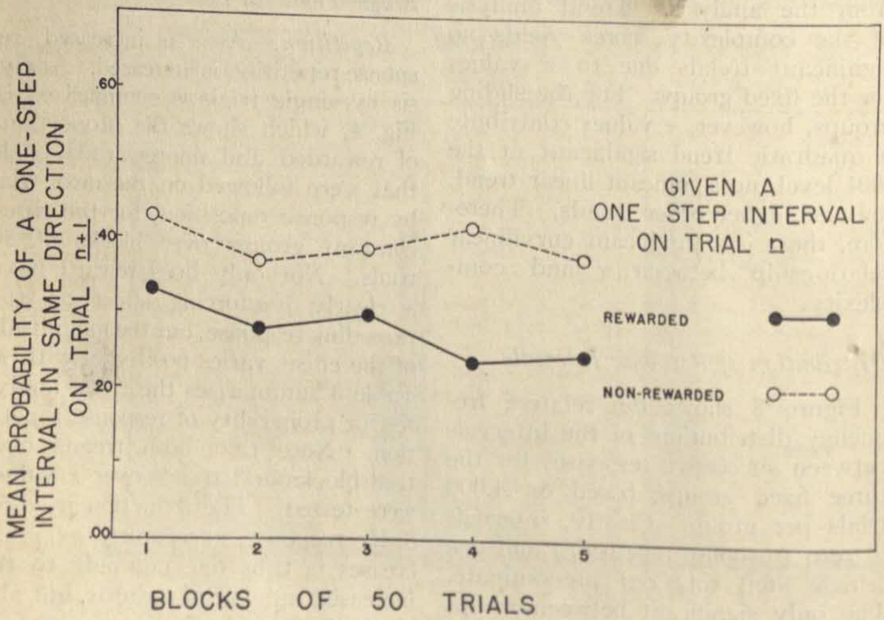


FIG. 5. Mean probability, given a one-step interval on Trial *n*, of a one-step interval in the same rotational direction on Trial *n* + 1 as a function of reward and nonreward on Trial *n*. (In the response sequence *i*, *j*, *k*, the interval, *i* − *j* is rewarded if and only if a reward occurs after Response *j* and before Response *k*.)

vided. In the case of response repetition, the overall density of reward interacts positively with the occurrence of a single reward to enhance its reinforcing power. Regardless of the overall reward density, however, a nonreward is more effective than a reward in producing a pattern of systematic rotation. Thus it might be argued that response repetition is an indication that *S* has a "solution," and one-step rotation is an indication that *S* has a systematic scanning strategy.[6] A qualification is the fact that at high densities of reward the scanning strategy may also be the solution.

Patterns of responding more subtle than repetition or rotation of responses occur most consistently when the uncertainty of outcome is greatest (i.e., when $\pi = .50$). This finding with respect to complexity suggests that when the inconsistency of feedback precludes both satisfactory convergence on a solution and orderly pursuit of a systematic scanning strategy, then *S* is forced to try more complex patterns, or simply to guess.

SUMMARY

Eighty *S*s were instructed that their responses on a circular array of pushbuttons determined which of two outcomes, reward or nonreward, would follow each response, but were in fact presented with preprogramed, random sequences of outcomes which differed systematically only in the relative frequency of reward. Various probabilities of reward were scheduled. Consistency of responding, defined as the repetition of any single button, was a positive, linear function of the probability of reward. The relative reinforcing power of a single reward for the next response was also a positive linear function of the probability of reward. When consistency due to orderly sequences of button selection was taken into account as well, the relation was positive, but curvilinear, consistency being greatest at high probabilities of reward, intermediate at low probabilities, and low at intermediate probabilities of reward.

Complexity of the response rules was greatest when the outcomes were most inconsistent (50% random reward), and least when the outcomes contained either high or low probabilities of reward. Complexity was also high when a period of 100% reward followed a relatively long period of nonreward.

It was concluded that superstitious response preferences and patterns are established as a function of the probability of reward, but that rotational sequences for the systematic testing of responses may also be established as orderly scanning rules, especially following nonreward. A distinction was made between collecting and using information—behaviors that are differentially influenced by the uncertainty of outcome and the probability of reward.

REFERENCES

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GUTHRIE, E. R., & HORTON, C. P. *Cats in a puzzle box.* New York: Rinehart, 1946.

JENKINS, H. M. An exploratory study of the development of concepts with noncontingent reinforcement. *Bell Lab. tech. Memo.*, 1959, No. 59–122–3.

LEWIS, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

MILLER, G. A., & FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt-Dryden, 1960.

MORSE, W. H., & SKINNER, B. F. A second type of superstition in the pigeon. *Amer. J. Psychol.*, 1957, **70**, 308–311.

PRIBRAM, K. H. On the neurology of thinking. *Behav. Sci.*, 1959, **4**, 265–287.

SHANNON, C. E. A mathematical theory of communication. *Bell Sys. tech. J.*, 1948, **27**, 623–656.

SKINNER, B. F. "Superstition" in the pigeon. *J. exp. Psychol.*, 1948, **38**, 168–172.

(Received May 5, 1961)

---

[6] Miller, Galanter, and Pribram (1960) present an excellent discussion of such plans for seeking information and finding solutions. Evidence concerning the distinction between organizing a search for information and organizing the use of information is offered by Pribram (1959) and by Jenkins (1959).

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 610–615

SACCHARINE CONCENTRATION AND DEPRIVATION AS DETERMINANTS OF INSTRUMENTAL AND CONSUMMATORY RESPONSE STRENGTHS[1]

HARRY L. SNYDER[2]

Johns Hopkins University

The importance of the consummatory response in the reinforcement process has been suggested by several authors. Sheffield, Roby, and Campbell (1954) have indicated that running speed increases with the average rate of ingestion of various sweet solutions. Similarly, Spence (1956) has suggested that a monotonic relationship should hold between running speed and the vigor of the consummatory response. If such consummatory response theories are to have any general utility, it follows that any variable or combination of variables which affects the strength of consummatory behavior should similarly affect the strength of the preceding running response. The present experiment examines this theoretical approach by analyzing concomitant variation in running and consummatory behavior as related to both deprivation and saccharine concentration. In addition, the data provide information about the manner in which drive and incentive combine to determine the strength of a running response.

METHOD

Subjects and Apparatus

The *S*s were 120 experimentally naive male rats of the Sprague-Dawley strain which were obtained from Sprague-Dawley, Incorporated, Madison, Wisconsin. The *S*s were 130 to 150 days old at the start of the experiment. They were housed in individual cages throughout the study.

The *S*s were trained in a 68-in. Plexiglas-covered, wooden alleyway. Guillotine doors separated the start and goal boxes from the running alley. The start box was 9 in. long and 6 in. wide, and the goal box was 11 in. long and 8 in. wide. The running alley was 4 in. wide. The entire alleyway was 9 in. high and was painted a flat black.

A microswitch, activated by raising the start-box door, and two photoelectric cells in circuit with Standard Electric timers automatically recorded starting latencies and running times. The photocells were located 1 in. and 45 in., respectively, past the start-box door.

A 3½-in. square piece of Plexiglas was centered in the wall at the far end of the goal box. A drinking tube was located in a ¾ × ⅜ in. vertical slot cut in the Plexiglas. The drinking tube consisted of a piece of ¼-in. brass tubing which contained a 2-mm. fluid hole.

With each lick upon the drinking tube, a pumping system delivered .004 ml. of fluid at the tip of the drinking tube. The pumping system, which is described in more detail elsewhere (Hulse, 1960), consisted of an infusion pump operated by an electronic relay circuit. The relay circuit also operated a counter and an Esterline-Angus pen recorder, Model AW, which indicated, respectively, the number of licks made upon the drinking tube, and the temporal distribution of licks. The paper speed of the recorder was 3 in. per sec. A stepping relay also activated a Standard Electric timer when *S* made the first lick of each trial. After the last lick on each trial, *E* stopped the timer. In this manner, the total time *S* took to make 150 licks was recorded.

Experimental Design

Four levels of food and water deprivation (2, 6, 12, and 22 hr. since last feeding) and three levels of saccharine concentration (0.25, 4.50, and 10.0 gm. saccharine added to 1 liter of water) were combined factorially to produce 12 experimental groups. Ten *S*s were randomly assigned to each group.

---

[1] This paper is based upon a dissertation submitted in partial fulfillment of the requirements for the PhD degree. The author is indebted to Stewart H. Hulse, director of the dissertation, for his helpful advice. This research was supported in part by a National Science Foundation Grant, G8712, to Stewart H. Hulse.

[2] Now at Colgate University.

Procedure

All experimental operations were carried out in an air-conditioned, humidity-controlled room under normal room illumination. These environmental conditions were maintained 24 hr. per day to avoid confounding of day-night cues with the times of running of the various groups. Temperature fluctuation throughout the experiment was less than 5° F.

Taming.—For 15 days, each *S* was fed a daily ration of 14 gm. dry Purina lab chow mixed with 29 ml. tap water to form a wet mash. This diet was maintained throughout the experiment. Following this diet-adaptation period, and continuing for 6 days, *S*s were handled freely and permitted to drink for 15 min. from a drinking tube containing the same concentration of saccharine that they would receive during training. Following each daily taming session, *S*s were returned to their living cages and fed at a time such that, for the following taming session, the appropriate deprivation state would be in effect. On the last day of taming, *S*s were handled but were not permitted to drink from the saccharine bottles.

Training.—Following taming, *S*s were given 35 training trials, 1 trial per day. Within each deprivation group, the order of running was randomized daily. On each trial, *S* was placed in the start box with the door lowered. After a delay of 2–3 sec., the door was raised and *S* was permitted to run to the goal box. After *S* entered the goal box, retracing was prevented by lowering the goal-box door. When the *S* had made 150 licks upon the drinking tube, it was removed to its living cage.

Extinction.—Following training, 10 extinction trials were given, 1 trial per day. All conditions were equivalent to the training conditions with the exception that no fluid was delivered through the drinking tube. All *S*s were permitted to lick upon the dry tube, however. Each *S* was removed from the apparatus 30 sec. after it entered the goal box.

Measures of Performance

Starting time (latency) was defined as the time which elapsed between *E*'s raising the start-box door and *S*'s breaking the first photocell beam. Running time was the time *S* took to run from the first to the second photocell. Running times were transformed to reciprocals to give running speeds.

Licking time was defined as the time which elapsed between the first and last licks each *S* made on each trial. The number of licks per trial (150) was divided by the licking time to give average rate of licking, in licks per second.

Esterline-Angus records were taken on Trials 33 and 34 of training. On each trial for each *S*, 4-sec. records were obtained following 10, 60, and 110 licks. Each 4-sec. record contained approximately 20–25 licks. The total time between the first and last licks was noted. From this total time, any time due to pauses between licks was subtracted. Such pauses were defined as intervals greater than twice the mean interlick interval for that record. The number of licks on the record was then divided by the remaining time (total time minus pauses) to give momentary rate of licking. Thus, while the average rate of licking is a mean value of licks per second throughout the entire session, the momentary rate of licking is a local rate at particular times within the session.

During extinction, licking behavior was measured by recording the number of licks each *S* made upon the dry tube during the 30-sec. goal confinement time.

RESULTS

Running Behavior

Training.—For the 35 training trials, starting latencies and running speeds for each *S* were averaged across seven blocks of 5 trials each. Analyses of variance were then used to assess the effects of deprivation (D), saccharine concentration (C), blocks of trials (Tr), and their interactions.

Starting latencies decreased significantly with increases in D ($F=3.33$, $df=3/107$, $P<.025$) and Tr ($F=17.61$, $df=6/642$, $P<.001$). No significant variation in latencies was obtained due to C ($P>.05$). None of the possible interactions was significant.

Running speeds increased significantly across Trials 1–35 ($F=12.34$,

TABLE 1

SELECTED RESULTS FROM RUNNING SPEED AND LICKING RATE ANALYSES OF VARIANCE

Source	*df*	Running Speed (Trials 16–35)		Average Rate (Trials 16–35)		Momentary Rate (Trials 33–34)	
		MS	*F*	*MS*	*F*	*MS*	*F*
Concentration (C)	2	27,627.58	21.88***	129.33	47.37***	1.08	2.84
Deprivation (D)	3	3,989.94	3.16*	11.09	4.06**	1.30	3.42*
C × D	6	368.40	—	3.73	1.37	1.53	4.02**
Between *S*s error	107[a]	1,262.77		2.73		0.38	

Note.—Within *S*s effects, whenever important, are discussed in text.
[a] One *S* was lost during the experiment and is treated statistically as a missing case.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

$df = 6/642$, $P < .001$). This increase with Tr was greater for increases in both C and D, as indicated by C × Tr and D × Tr interactions ($F = 7.62$, $df = 12/642$, $P < .001$ and $F = 1.92$, $df = 18/642$, $P < .05$, respectively). The C × D interaction was not significant.

In order to further evaluate the effects of C and D, a separate analysis was used for Trials 16–35. The results of this analysis are presented in Table 1. For Trials 16–35, increases in either C or D led to increased running speeds, as shown in Fig. 1 and 2. In addition, the C × D interaction was not significant ($F < 1.00$), which indicates that the effects of C and D upon running speed are independent. Neither Tr nor

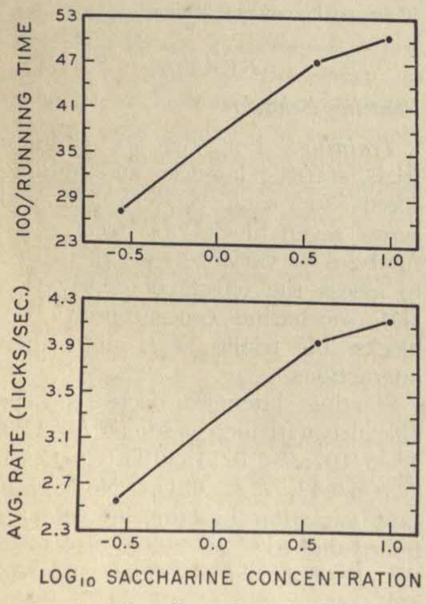


FIG. 1. The effect of saccharine concentration upon running speed and average rate of licking, Trials 16–35.

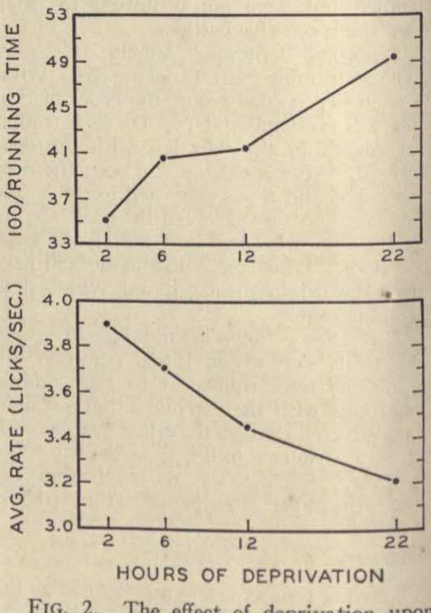


FIG. 2. The effect of deprivation upon running speed and average rate of licking, Trials 16–35.

any of the interactions involving Tr were significant for Trials 16–35.

Extinction.—Starting latencies and running speeds for the 10 extinction trials were analyzed with respect to individual trials rather than blocks of trials.

Increases in both C and D led to faster running speeds during extinction ($F=17.28$, $df=2/107$, $P<.001$, and $F=5.72$, $df=3/107$, $P<.01$, respectively). The interaction between C and D was not significant.

Significant C × Tr and D × Tr effects were obtained for the 10 extinction trials ($F=2.74$, $df=18/963$, and $F=2.86$, $df=27/963$, respectively, both Ps $< .001$). The C × Tr interaction indicates that, for the three C groups, the speed curves converge over the 10 trials. However, the D × Tr interaction presents a different picture. In this case, inspection of the group means indicates that the 22-hr. deprived *S*s show very little, if any, extinction, while the other three D groups extinguish relatively rapidly.

Consummatory Behavior

Average rate of licking.—The average rate of licking for Trials 16–35 (Fig. 1 and 2) increases with saccharine concentration ($P<.001$), but *decreases* with deprivation ($P<.01$). The C × D interaction was not significant (Table 1). Neither Trials nor any of the interactions involving Trials were significant.

Momentary rate of licking.—Momentary rates for Trials 33 and 34 varied significantly with both D and the C × D interaction, as shown in Table 1. The significant D effect was nonmonotonic and is probably due to the C × D interaction, which is illustrated in Fig. 3. Figure 3 shows that the momentary rate of licking increases as a function of saccharine concentration for *S*s under relatively low deprivation (2 and 6 hr.), remains nearly constant for *S*s under moderate deprivation (12 hr.), and decreases as a function of concentration for highly-deprived *S*s (22 hr.). For convenience, the 2- and 6-hr. groups have been combined in Fig. 3 since they did not differ significantly.

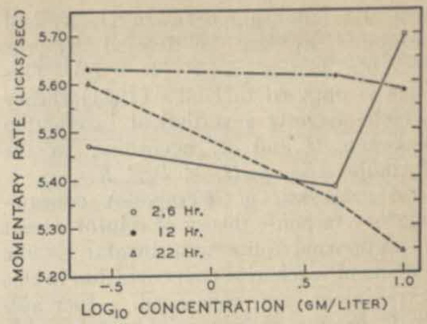


FIG. 3. Momentary rates of licking for combinations of deprivation and saccharine concentration, Trials 33 and 34.

Momentary rates of licking also decreased as licking within a session progressed ($F=1010.67$, $df=2/211$, $P<.001$). Moreover, this decrease with number of licks was greater for higher concentrations as indicated by a small but highly significant C × Number-of-Licks interaction ($F=3.67$, $df=4/211$, $P<.01$).

Extinction of licking behavior.—The number of licks made upon the dry tube during the 10 extinction trials did not vary significantly with either C, D, or C × D.

DISCUSSION

Running behavior.—The present results agree with previous experiments (Reynolds & Pavlik, 1960; Weiss, 1960) which have shown that deprivation and incentive have independent effects upon strength of response. The results thereby add to the evidence supporting Spence's (1956) notion that the theoretical relationship between D, K, and response strength is defined by the formula: $_sE_R = {_sH_R} \times (D + K)$. This idea is opposed to Hull's (1951) theory which suggests a statistical interaction between D and K, according to the formula: $_sE_R = {_sH_R} \times D \times K$.

Average rate of licking.—A consummatory response theory of reinforcement is supported quite well insofar as the effects of saccharine concentration in this experiment are concerned. Increases in C produce increases in both running speed and average rate of ingestion (see Fig. 1). However, increases in deprivation produce increases in running speed but *decreases* in average rate of ingestion (see Fig. 2). This decrease is not affected by saccharine concentration as indicated by the nonsignificant C × D interaction. Hence, the utility of a consummatory response theory of reinforcement may well be limited to variables dealing with the nature of the reinforcer, per se, such as sweetness or volume, but apparently does not extend to variables dealing with the deprivation state of the organism. Quite obviously, these results place a severe restriction upon the general usefulness of such theories.

Momentary rate of licking.—When they lick, rats lick faster for higher concentrations if they are slightly deprived, but lick faster for lower concentrations if they are greatly deprived (see Fig. 3). It would appear, as a first approximation, that this interaction is related to sensory events. Specifically, it may be the case that the effect of deprivation is to change the threshold for sweet stimulation and thus cause the taste receptor stimulation for saccharine solutions to vary complexly with the deprivation state of the rat. Indeed, there are data to indicate that both threshold and preference for sweet solutions are altered by deprivation (Bacon, Snyder, & Hulse, 1962; Campbell, 1958).

In a related manner, licking behavior within a session may also be determined by sensory events. The momentary rate decreases as a function of the number of licks, and this decrease with number of licks is greater for both higher concentrations and greater volumes per lick (Snyder & Hulse, 1961). Thus, when the rate of ingestion is directly contingent upon an instrumental rate of responding, there is some regulatory mechanism for ingestion which reduces this rate of responding, and the reduction in rate is greater for larger volumes or higher concentrations. Whatever the locus of the mechanism, it operates both for licking, as shown here, and for bar pressing, as indicated by Collier and Siskel (1959).

SUMMARY

Three levels of saccharine concentration (0.25, 4.50, and 10.0 gm/l) and four levels of food and water deprivation (2, 6, 12, and 22 hr.) were combined factorially to produce 12 experimental groups. Ten male albino rats per group were given 35 training and 10 extinction trials, 1 trial per day, in a straight alleyway. On each trial, *S*s were permitted to make 150 licks upon a drinking tube which supplied the saccharine solution. Running speed and consummatory behavior were measured.

The results were: (*a*) Running speed increased with both deprivation (D) and concentration (C). The C × D interaction was not significant. (*b*) The average rate of licking (licks/sec measured over the entire reinforcement session in the goal box) increased with C but *decreased* with D. (*c*) The momentary rate of licking (local rate within bursts of licking) varied complexly with the C × D interaction. Slightly-deprived rats lick faster for higher saccharine concentrations, but greatly-deprived rats lick faster for lower concentrations.

The additive relationship between C and D for running speeds was interpreted as supporting Spence's theoretical formulation, $_sE_R = {_sH_R} \times (D+K)$. Although the monotonic relationship between running speed and average rate of licking for the C variable firmly supports a consummatory response theory of reinforcement, the negative relationship between these response measures for the D variable clearly indicates a severe limitation to such a theory.

REFERENCES

BACON, W. E., SNYDER, H. L., & HULSE, S. H. Saccharine preference in satiated and deprived rats. *J. comp. physiol. Psychol.*, 1962, **55**, 112–114.

Campbell, B. A. Absolute and relative sucrose preference thresholds for hungry and satiated rats. *J. comp. physiol. Psychol.*, 1958, **51**, 795–800.

Collier, G., & Siskel, M., Jr. Performance as a joint function of amount of reinforcement and inter-reinforcement interval. *J. exp. Psychol.*, 1959, **57**, 115–120.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1951.

Hulse, S. H., Jr. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, **3**, 1–3.

Reynolds, W. F., & Pavlik, W. B. Running speed as a function of deprivation period and reward magnitude. *J. comp. physiol. Psychol.*, 1960, **53**, 615–618.

Sheffield, F. D., Roby, T. B., & Campbell, B. A. Drive reduction versus consummatory behavior as determinants of reinforcement. *J. comp. physiol. Psychol.*, 1954, **47**, 349–354.

Snyder, H. L., & Hulse, S. H. Effects of volume of reinforcement and number of consummatory responses on licking and running behavior. *J. exp. Psychol.*, 1961, **61**, 474–479.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Weiss, R. F. Deprivation and reward magnitude effects on speed throughout the goal gradient. *J. exp. Psychol.*, 1960, **60**, 384–390.

(Received May 10, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 616–620

CONCEPT IDENTIFICATION AS A FUNCTION OF OBVIOUSNESS OF RELEVANT AND IRRELEVANT INFORMATION[1]

E. JAMES ARCHER

University of Wisconsin

In a number of previous studies (e.g., Archer, 1954; Archer, Bourne, & Brown, 1955; Brown & Archer, 1956; Bulgarella & Archer, 1962) a binary-digit scale ($\log_2$) was used to describe the amount of information in stimuli used in concept identification tasks. In addition the information could be qualitatively described as relevant, irrelevant, or redundant. In all of the studies 1 bit of information from one dimension was treated as equivalent to an equal amount of information from any other dimension. As a first approximation, such an application of information analysis is an oversimplification. For example, the Brown and Archer (1956) study indicated that information concerning the physical location of the stimulus apparently was less obvious to *S* than information concerning form or color. Concepts having the spatial location of the stimulus relevant proved to be more difficult than others not having spatial location relevant. It seems reasonable to suppose that the obviousness of information is a manipulable variable and that such a characteristic of the information affects concept identification. Furthermore, at least one simple prediction can be made concerning the relationship between obviousness and relevance: there should be a significant interaction between the two variables, i.e., if the relevant information were also obvious, the concept should be easier to discover than if the information were less obvious. On the other hand, if the irrelevant information were obvious, the concept should be more difficult to discover than if the information were less obvious. The first part of this hypothesis is self-evident; the latter part assumes that when *S*'s attention is drawn to irrelevant information, he will attempt a solution involving information which should not be used, and when the irrelevant information is less obvious *S* is better able to ignore the information which should be ignored.

In order to assess the generality of this prediction, it seemed desirable to study two distinctly different dimensions. Because of the simplicity and accuracy of control, the dimensions of *form* and *size* were selected.

PROCEDURE

Subjects.—A total of 128 students (64 men and 64 women) from introductory psychology courses at the University of Wisconsin served individually as *S*s in this study.

Conditions.—As may be seen from Table 1 there were 16 combinations of stimulus conditions of Manipulated Dimension, Obviousness, Relevance (of Manipulated Dimension), and Level of Irrelevant Information arranged in a 2^4 orthogonal design.

Since the effect of sex of *S* was being studied, 4 men and 4 women, otherwise randomly selected, were assigned to each of the 16 combinations of stimulus conditions.

---

[1] This study was supported by a National Science Foundation grant (G-1691). The author is indebted to Matilda E. Levine who made the stimulus materials and gathered the data of the main study.

This report was prepared while the author was a National Science Foundation Senior Postdoctoral Fellow at the Applied Psychology Research Unit, Cambridge, England.

Apparatus.—The stimuli were back-projected onto an 8 × 10 in. opal-glass screen by a Dunning Animatic 16-mm. strip-film projector. The screen was mounted in the wall of the booth above *S*'s control console. The booth was made of sound-absorbent material 4 × 4 × 8 ft. and painted flat black inside. The control console had four equally spaced lever-action switches. Above each switch was a neon glow lamp.

Task.—The *S* was told that he would see different patterns and that his task was to assign each of the patterns to one of four categories. He was to indicate his assignment by pressing one of the switches before him. When a switch was operated, one of the four glow lamps would immediately light over the switch which should have been pressed. The *S* was urged to press one and only one switch firmly and then release it. The glow lamp remained lit for 2 sec. and then 1 sec. later the strip-film projector advanced to the next pattern.

The task was partially self-paced since *S* could take as long as he wished to decide which switch to depress, but after responding, the next pattern automatically appeared in 3 sec.

The *S* was told that all of the patterns in a particular category would have something in common and that he would continue at the task until he had made 32 consecutively correct responses.[2]

RESULTS

Time to criterion.—Because of heterogeneity of variance the time-to-criterion scores were logarithmically transformed and the total variance of these latter scores was then analyzed. Since this was a 2^5 orthogonal design replicated four times, each main effect and interaction had 1 *df* and the error term had 96 *df*.

As expected the amount of irrelevant information was an effective variable. The mean time to criterion was 10.25 min. and 17.98 min. for 1 and 3 bits irrelevant, respectively ($F = 30.429$, $P < .001$).

TABLE 1

DESIGN OF THE EXPERIMENT

Manipulated Dimension	Obviousness	Relevance	Stimulus Dimensions Varied[a]		
			Relevant	Irrel. (1)	Irrel. (3)
Size	High (1 in. and ¾ in.)	Relevant	size number	color	color shade horizontal
		Irrel.	color number	size	size shade horizontal
	Low (1 in. and ⅞ in.)	Relevant	size number	color	color shade horizontal
		Irrel.	color number	size	size shade horizontal
Form	High (90° and 84°)	Relevant	form dot	shade	shade number horizontal
		Irrel.	shade dot	form	form number horizontal
	Low (90° and 87°)	Relevant	form dot	shade	shade number horizontal
		Irrel.	shade dot	form	form number horizontal

Note.—This design was duplicated for 64 men and 64 women students. Four *S*s of each sex served in the 16 combinations of relevant and irrelevant information.

[a] The levels of the dimensions were as follows: Size (large or small); Number (one or two identical figures); Color (red or green); Shade (light or dark); Horizontal (figure appeared on left or right of screen); Form (square or parallelogram—87° or 84° opposite acute angles); Dot (presence or absence of a ¼-in. white dot centrally located on figure).

There was also a difference in difficulty depending upon the manipulated dimension. The problems involving variations in size were easier (9.36 min.) than those involving variations in form (18.87 min.), $F = 31.441$, $P < .001$. None of the other main effects was significant.

As shown in Table 2 the predicted interaction of Relevance × Obviousness was highly significant. The meaning seems quite clear; if the information (and in this case we mean the manipulated information of either size or form) is obvious and relevant,

---

[2] The choice of the different sizes and forms was determined by a pilot study which assured 100% discriminability of the stimuli when *S*s were asked to sort for size or form. Unless the stimuli were 100% discriminable, the *S*s would not be able to solve the concept identification task.

TABLE 2

MEAN TIME TO CRITERION (MIN.) AS A FUNCTION OF OBVIOUSNESS AND RELEVANCE OF INFORMATION

Obviousness	Relevance	
	Relevant	Irrelevant
High	12.7	15.6
Low	16.5	11.7

Note.—Each mean is based on an N of 32. Because of heterogeneity of variance, the analysis was performed on log-transformed scores; $F = 7.172$, $df = 1/96$, $P < .01$.

then the criterion is achieved fairly easily. If, however, the information is equally obvious but irrelevant, then the task proves fairly difficult. Just the opposite obtains when the information is not obvious but is relevant; the task is very difficult. Finally, when the equally unobtrusive information is irrelevant, then the task is easiest of all.

Another first-order interaction which was significant was that of Relevance × Manipulated Dimension. This relationship is shown in Table 3.

Clearly the effects of the two manipulated dimensions are different. When form was relevant the task was quite difficult; and when form was irrelevant, the task became very difficult. Exactly the opposite occurred with the variation in relevance of size; when relevant the task was fairly easy and when irrelevant it became even easier. This perplexing differential effect of the two manipulated dimensions might have remained unresolved except for the results shown in Table 4.

TABLE 3

MEAN TIME TO CRITERION (MIN.) AS A FUNCTION OF MANIPULATED DIMENSION AND ITS RELEVANCE

Manipulated Dimension	Relevance	
	Relevant	Irrelevant
Size	11.9	6.8
Form	17.2	20.5

Note.—Each mean is based on an N of 32. Because of heterogeneity of variance, the analysis was performed on log-transformed scores; $F = 14.418$, $df = 1/96$, $P < .01$.

Here we see an unexpected significant Sex × Manipulated Dimension × Relevance interaction. As is evident, the two sexes do not behave differently when size was relevant or irrelevant. However, when form was relevant, men found the task fairly easy; and when form was irrelevant the men found the task very difficult. Just the opposite effect was shown by the women. When form was relevant, the women found the task quite difficult but when form was irrelevant, they found the task much easier.

TABLE 4

MEAN TIME TO CRITERION (MIN.) AS A FUNCTION OF SEX OF S, MANIPULATED DIMENSION, AND RELEVANCE OF THE DIMENSION

Sex of S	Manipulated Dimension	Relevance	
		Relevant	Irrelevant
Men	Form	12.9	24.4
	Size	12.7	8.1
Women	Form	21.6	16.6
	Size	11.1	5.5

Note.—Each mean is based on an N of 16. Because of heterogeneity of variance, the analysis was performed on log-transformed scores; $F = 4.355$, $df = 1/96$, $P < .05$.

None of the other main effects or interactions was significant.

Errors to criterion.—A different, though related ($r = .93$, $df = 126$, $P_{r=0} < .001$), measure is that of errors to criterion. Again, because of heterogeneity of variance a logarithmic transformation was applied to these scores. An analysis of variance yielded results similar to the time analysis; the same main effects and interactons were significant except for the Sex × Manipulated Dimension × Relevance interaction which just failed of the .05 level of confidence.

DISCUSSION

Two conclusions seem appropriate for this study. First, while an information analysis of the stimuli used in a concept identification task provides a starting point for predicting the difficulty of the task for *S*, the performance can be greatly modified by a secondary stimulus quality such as *obviousness* of information.

As shown by the present study, the optimum conditions to enable *S* to identify a concept will be when the obviousness of the relevant information is maximized and the obviousness of irrelevant information is minimized. The ultimate, of course, of this latter condition is when the difference between (or among) the levels is zero and then the irrelevant information no longer exists.

One source of variation usually described as *individual differences* may be partly due to the perceptual biases of the individual *S* which alter the obviousness of the stimulus differences. In some cases, *S* may perceive and attend to stimulus differences which are relevant and will thereby perform well. In other cases, *S*s who perceive and attend to the irrelevant information will, of course, perform poorly. As this study shows, we can vary the probability of *S* attending to the relevant and irrelevant information.

Perhaps the most intriguing aspect of the present study is the observation of the sex difference, not as a main effect, but as the Sex × Manipulated Dimension × Relevance interaction. Since it seemed unlikely that this was a biologically determined phenomenon, a perceptual, albeit post hoc, explanation was sought.

The same stimuli which had been used in the most complex form-variation condition were shown to 40 men and 40 women students and they were asked to describe what they saw. Aside from number, color, and white dot, men usually described the stimuli as squares and parallelograms or even trapezoids. Women, on the other hand, much more frequently described the stimuli as squares and nonsquares, or not squares, or even tippy squares. While recognizing the dangers of a post hoc explanation and the especial hazard of generalizing from one sample to another sample, it is hard to resist the hypothesis that the obviousness of a level within a dimension can be modified by *S*'s labeling ability. If the men in the experimental study could label the stimuli with distinctly different names, e.g., square and parallelogram or even trapezoid, then the obviousness of the form information would be enhanced (relative to not having a label for the different levels) and if the information about form were relevant, then the task would be easier than if the information were irrelevant. Furthermore, just the opposite would occur if the women in the experimental study did not have distinctly different labels, e.g., square and nonsquare; they would find the task very easy when form was irrelevant and very difficult when form was relevant.

The results suggest the way inner speech might modify problem solving behavior. It seems reasonable to predict that any verbal pretraining which successfully modifies inner speech will thereby affect performance in problem solving tasks like concept identification.

SUMMARY

A 2^5 orthogonal design replicated four times examined the effect of obviousness of stimulus differences upon concept identification when the manipulated information was either relevant or irrelevant. It had been predicted that there would be an interaction between relevance and obviousness so that performance would be facilitated if the relevant information were obvious, and performance would be impeded if the relevant information were less obvious; and the reverse would obtain for the irrelevant information. The prediction was supported.

An unexpected result indicated that men and women perceived the two forms of squares and parallelograms differently. Men found the task easier when *form* was relevant rather than irrelevant; and the opposite obtained for women. A post hoc explanation was offered; the differential performance might have been due to the unequal availability of distinctive labels for the forms by the two sexes.

REFERENCES

ARCHER, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

BULGARELLA, R. G., & ARCHER, E. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *J. exp. Psychol.*, 1962, **63**, 254–257.

(Received May 11, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 6, 621–626

RECIPROCAL AUGMENTATION OF GENERALIZATION AND ANXIETY[1]

SARNOFF A. MEDNICK AND CYNTHIA WILD
University of Michigan *Yale University*

As a part of a general theory of schizophrenic behavior it has been suggested that the core of the acute breakdown in schizophrenia involves an interaction between learned fear and generalization (Mednick, 1958). It is hypothesized that under certain conditions these two factors could reciprocally augment each other. The hypothesis of reciprocal augmentation suggests that if stimuli which elicit fear reactions were repeatedly presented to *S* his general level of arousal would tend to rise. As a consequence, his responsiveness to fear producing generalization stimuli would tend to be augmented. If such generalization stimuli were to continue to evoke reactions from him, his general level of arousal would be again augmented which would again produce an increase in responsiveness to fear producing generalization stimuli. The key to the uniqueness of this process is the fact that stimuli which elicit fear responses can bring about increases in level of arousal.

The process may be conceptualized in detail as follows. Let GS_1, GS_2, GS_3, and GS_4 be a series of generalization stimuli which elicit fear responses from *S*; let A be *S*'s arousal level at the inception of the series of stimuli. When GS_1 is presented, *S* will react with fear. Since learned fear contributes to arousal (Malmo, 1957), the new momentary arousal level will be $A + \Delta 1$. One important consequence of this increment in arousal level will be the raising of the height and breadth of the gradient of generalization responsiveness (Mednick & Freedman, 1960). This augmentation of generalization will be temporary, decreasing as $\Delta 1$, the fear response to GS_1, decays. However, if GS_2 is presented soon enough after GS_1 it will stimulate *S* while he is in a state of heightened arousal and will consequently elicit a relatively larger fear response than GS_1 had elicited. This response will then send the arousal level up to $A + \Delta 1 + \Delta 2$ (less some quantity ascribable to the decay of $\Delta 1$). This temporary rise in arousal level will in turn again augment generalization responsiveness. Now, if GS_3 is presented soon enough after GS_2, it will elicit a relatively larger response, add to total arousal level, augment the response to GS_4 . . . and so on.

One distinctive and testable prediction that the reciprocal augmentation hypothesis generates suggests that a learned fear response (and generalized responses) would extinguish more slowly under massed practice than distributed practice. The longer the time between trials the greater the opportunity for the fear response from the preceding trial to decay and drop the arousal level to the resting state; if trials follow each other rapidly, the arousal level and generalization responsiveness level should tend to rise. Thus, arousal level and generalization would be interacting so as to retard extinction (or even overcome the effects of

---

[1] This study was carried out while the senior author was at Harvard University and was supported by grants from the National Association for Mental Health and National Science Foundation (Grant No. 3855). Elizabeth Brunette helped carry out a similar preliminary study using the dimension of pitch generalization. Her help with the apparatus of this study is greatly appreciated.

extinction and cause increasing responsiveness for some limited time).

Miller (1951) has hinted at the possibility of such a "bootstraps" phenomenon in anxiety. He questioned whether (in accordance with the results of studies using instrumental responses) massed trials would lead to the extinction of a learned fear response more rapidly than spaced trials. He also raised the possibility "that other effects, such as perseveration or summation of fear might reverse the (extinction) curve somewhat . . ." (p. 450).

In this study, it is predicted that the rate of extinction of a learned fear response and generalized responses will vary inversely with the speed of intertrial interval during extinction.

METHOD

The *S*s were trained to evidence a PGR reaction to a crucial word and tested with words which elicited this crucial word as an associate. In a factorial design, two intertrial intervals were used in conditioning and extinction.

Subjects.—The *S*s were 60 paid volunteer Harvard Summer School students. In order of appearance, they were assigned to four groups of 15 *S*s each, with the restriction that each subgroup was approximately evenly divided as to sex. The *S*s were informed that they would receive electric shocks.

Apparatus.—Galvanic measures were made with an apparatus suggested and described by Haggard and Gerbrands (1947). Stoelting finger electrodes were used with Cambridge electrode jelly.

The verbal stimulus materials were presented to *S* via earphones by means of a tape recorder. A measure of the intensity of the verbal stimuli revealed no differences between stimuli. A Harvard inductorium equipped with 1½-v. batteries delivered a shock which was judged quite painful but bearable by preliminary *S*s. This shock level was maintained for all *S*s, the batteries being periodically replenished.

Verbal stimulus materials.—The CS and mediated generalization stimuli (GS) were chosen by a method used by Martha Mednick (1957). They fit a paradigm of mediated generalization in which a CR is first trained to a CS. When a GS occurs which elicits the CS, the CR is in turn elicited. In this study the GS-CS connection was taken from the Russell and Jenkins (1954) revision of the norms of the Kent-Rosanoff Word Association Test. The auditory presentation of the word LIGHT was chosen as the CS. The GSs were the words LAMP and SOFT, both of which elicit LIGHT as a response (with probabilities of .63 and .0087, respectively). The word SQUARE was chosen as a neutral stimulus; it did not elicit LIGHT as a response in the revised norms. Thus, it was assumed that, after conditioning a PGR to LIGHT, when LAMP was presented it would elicit the implicit association LIGHT which in turn would be the stimulus for the conditioned PGR. It has been shown that this response generalizes from the auditory presentation to the implicit thought response (Noble, 1950).

During conditioning *S* heard, via earphones, seven neutral words and seven presentations of the CS. One-half second after the onset of five of the CS presentations *S* received a shock to the left wrist. Partial reinforcement during conditioning was used to prolong extinction and provide tests of conditioning level; the third and seventh presentations of the CS were not followed by the UCS.

The extinction series consisted of the CS, two generalization stimuli, and the neutral stimulus. These were presented four times each for a total of 16 extinction trials. The order of presentation in the extinction series was counterbalanced to avoid order effects.

Conditioning and extinction.—There were four groups. Group 8-20 had an 8-sec. intertrial interval in conditioning and a 20-sec. intertrial interval during extinction. These intertrial intervals were measured from the onset of CS in one trial to the onset of CS in the next trial. The other groups were Groups 8-8, 20-8, and 20-20.

The experiment was conducted in a quiet, dry, and cool basement room. The earphones limited the possibility of distracting noises. The *S* was shown the apparatus, seated in a broad-armed chair facing a wall, and had electrodes and earphones attached. It was explained that he would receive some electric shocks and hear some words; he was asked to remain quiet since physiological reactions were being taken. After an adaptation period in which *S* reached a steady PGR baseline, the conditioning series was begun. The generalization-extinction series followed the conditioning series with no interruption.

RESULTS

Raw ohms resistance data were obtained from the records by measuring the heights of the momentary individual galvanic deflections and referring to a previously prepared conversion chart (Haggard & Gerbrands, 1947). Ohms resistance (r) was then transformed into micromhos (M) by use of the formula $M = 10^6 \frac{1}{r}$. Due to the skewed nature of the data, medians and appropriate nonparametric statistical analyses were used in almost all instances.

Conditioning intertrial interval.—The two CS trials not followed by shock during conditioning were included as a measure of extent of conditioning. However, for the groups having a conditioning intertrial interval of 8 sec. (8-8, 8-20), only the second of these trials was scoreable. The 8-sec. interval was not sufficient for *S* to recover from his UCR to the last UCS presentation. However, the one trial which was almost always scoreable was the unreinforced presentation of the CS at the conclusion of conditioning. This is an indicator of final conditioning level reached. The median response of the 20-20 and 20-8 groups was .0431 M; the median response of the 8-20 and 8-8 groups was .0179 M. The difference evaluated by the Mann-Whitney *U* test was significant ($z = 4.27, P < .001$), indicating that the 20-sec. intertrial interval in conditioning resulted in greater responsiveness to the CS at the end of conditioning.

Another way of evaluating the effect of the conditioning intertrial interval is to compare the resultant level of generalization responsiveness and ease of extinction. There were 16 generalization-extinction trials composed of the CS, two GSs, and the neutral stimulus, each presented four

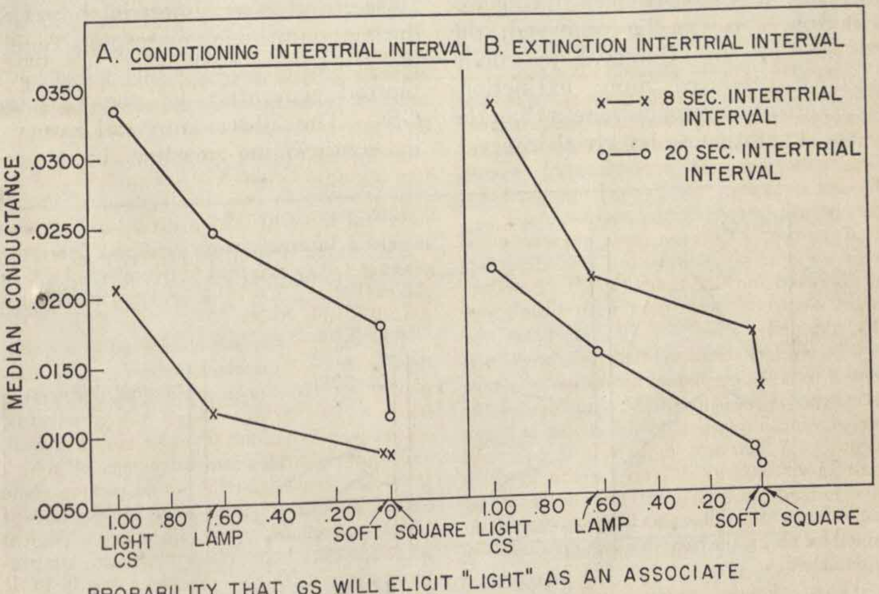


FIG. 1. Effect of intertrial interval during conditioning (A) and extinction (B) on generalization responsiveness.

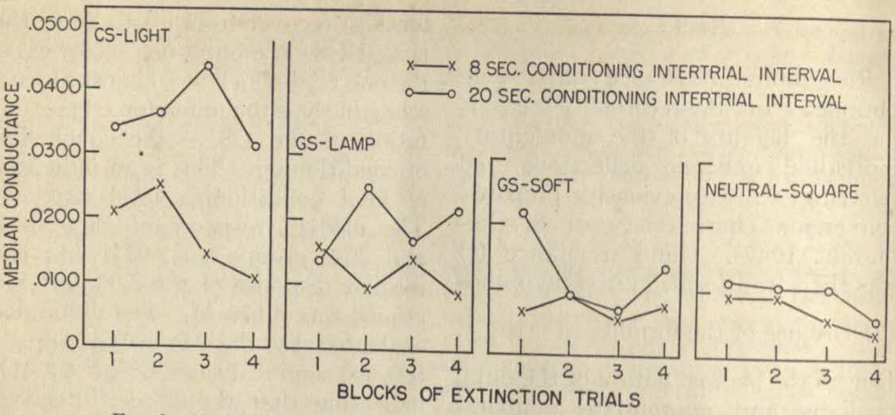


FIG. 2. Extinction of PGR to CS and GSs as a function of intertrial interval during conditioning.

times. By taking the median response to each GS and arranging the GSs in order of the probability that they would elicit the CS as an associated response, gradients of mediated generalization may be constructed for each group as in Fig. 1A. Figure 1A shows the effect of the intertrial interval during conditioning on generalization responsiveness during extinction. As can be observed, the 20-sec. intertrial interval in conditioning produced more extinction-generalization responsiveness than the 8-sec. conditioning intertrial interval.

In order to obtain a total extinction-generalization score the amplitudes of responses to the test words were summed for each *S*. The differences between the 8-sec. and 20-sec. conditioning intertrial interval on this score proved significant in a Mann-Whitney *U* test ($z = 2.91$, $P < .001$).

Figure 2 compares the effect of the 20-sec. and 8-sec. intertrial intervals during conditioning on responsiveness in extinction. Extinction curves are plotted separately for the CS and GSs. The 20-sec. intertrial interval in conditioning resulted in slower

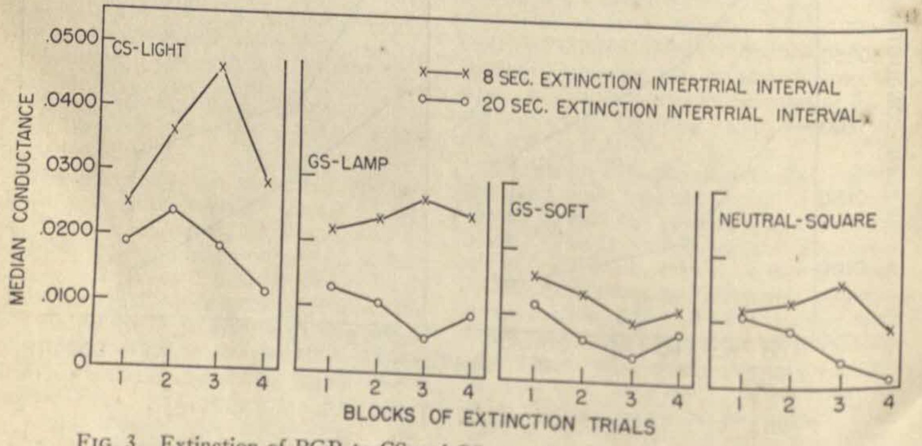


FIG. 3. Extinction of PGR to CS and GSs as a function of intertrial interval during extinction.

extinction of responsiveness to all of the stimuli.

Extinction intertrial interval.—Figure 1B shows the effect of the intertrial interval during extinction on generalization responsiveness. The 8-sec. intertrial interval during extinction resulted in more generalization responsiveness than the 20-sec. intertrial interval. The difference in total generalization score evaluated by the Mann-Whitney *U* test proved highly significant ($z=3.21$, $P<.001$).

Figure 3 compares the effect of the 20-sec. and 8-sec. intertrial intervals during extinction on responsiveness during extinction. Extinction curves are plotted separately for the CS and GSs. The 8-sec. intertrial interval in extinction resulted in slower extinction than the 20-sec. extinction intertrial interval. For the CS and GS, LAMP, there is no evidence of any lessening of responsiveness at the conclusion of the extinction series; if anything, the amplitude of responses seems to increase as "extinction" proceeds.

Inspection of the extinction curves for the four subgroups did not suggest that there had been any interactive effect of conditioning and extinction intertrial intervals. The curves for Groups 20-8 and 8-8 were essentially parallel as were the curves of Groups 8-20 and 20-20. An analysis of variance supported this impression; the interaction term was not significant ($F = .12$, $df = 1/56$). The main effects were significant.

DISCUSSION

Spaced conditioning trials resulted in greater amplitude of response at the conclusion of conditioning, more generalization, and more resistance to extinction than massed conditioning trials. While this finding is in accord with most previous research and theory (Hull, 1943; Rohrer, 1947) it is somewhat surprising in this study. The effectiveness of the distributed practice had to be demonstrated in the face of the presumably higher arousal level under the 8-sec. conditioning intertrial interval. However, the conditioning amplitude result should be interpreted with caution since it is based on data from only one trial.

Massed extinction trials with the CS and GSs resulted in greater generalization and more resistance to extinction than spaced extinction trials. A similar result for resistance to extinction has been reported by Sheffield (1950). She suggested that the massed trials group suffered greater frustration in extinction leading to greater drive leading to greater resistance to extinction. The results of this study seem to extend the generality of her interpretation. That is, any aspect of the extinction procedure that will increase arousal will tend to slow extinction. The more arousal provoked, the slower the extinction process. In the present study special advantage was taken of the arousal properties of the fear response. In the massed trials condition, fear eliciting stimulation (CS and GSs) was administered before the fear from the preceding trial could decay, making for stronger fear responses, greater accumulation of arousal, and slower extinction. These results are even more impressive in view of the fact that extinction of response to generalization stimuli is usually very rapid. Two recent studies have reported similar findings. Both used a noxious UCS and found more resistance to extinction with massed trials in extinction; both human and animal *S*s were studied. Murphy and Miller (1956) used an electric shock UCS in a modified Warner box; Howat and Grant (1958) studied eyelid conditioning. Both found faster extinction under spaced extinction trials.

The relationship between the reciprocal augmentation hypothesis and the acute schizophrenic process is discussed elsewhere (Mednick, 1958).

SUMMARY

A psychogalvanic reaction was trained to the word LIGHT under conditions of massed and distributed trials. In a factorial design the CR to the CS, LIGHT and mediated generalization stimuli, LAMP, SOFT, and SQUARE, were extinguished with 8-sec. and 20-sec. intertrial intervals.

Spaced trials in conditioning produced greater amplitude of CR, greater generalization, and more resistance to extinction. Massed trials in extinction led to greater generalization and slower extinction. An interpretation of the results is offered in terms of reciprocal augmentation of anxiety and generalization, a phenomenon thought to have relevance for psychopathology.

REFERENCES

HAGGARD, E. A., & GERBRANDS, R. An apparatus for the measurement of continuous changes in palmar skin resistance. *J. exp. Psychol.*, 1947, **37**, 92–98.

HOWAT, M. G., & GRANT, D. A. Influence of intertrial interval during extinction on spontaneous recovery of conditioned eyelid responses. *J. exp. Psychol.*, 1958, **56**, 11–15.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–287.

MEDNICK, M. T. Mediated generalization and the incubation effect as a function of manifest anxiety. *J. abnorm. soc. Psychol.*, 1957, **55**, 315–321.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55**, 316–327.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

MURPHY, J. V., & MILLER, R. E. Spaced and massed practice with a methodological consideration of avoidance conditioning. *J. exp. Psychol.*, 1956, **52**, 77–81.

NOBLE, C. E. Conditioned generalization of the GSR to a subvocal stimulus. *J. exp. Psychol.*, 1950, **40**, 15–25.

ROHRER, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, **37**, 473–493.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. 8-ONR-66216)

SHEFFIELD, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.

(Received May 18, 1961)

LIBRARY